

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:40:28 ; Search time 0.910506 Seconds  
(without alignments)  
1479.047 Million cell updates/sec

Title: US-09-662-293-1  
Perfect score: 78  
Sequence: 1 SIKRDHNDYSKNPM 14

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: PIR.78:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	57.7	787	2	PC1232
2	45	57.7	1409	1	OFFFCP
3	42	53.8	840	2	probable sulfate p
4	42	53.8	900	2	galactosamine-cont
5	41	52.6	906	2	preproteins translo
6	40	51.3	202	2	T36368
7	40	51.3	383	2	E70184
8	40	51.3	548	2	T36881
9	39.5	50.6	759	2	trans-Golgi membra
10	39	50.0	367	2	P95159
11	39	50.0	367	2	phenanthrene dehydro
12	39	50.0	383	2	D90528
13	39	50.0	412	2	C96816
14	39	50.0	444	2	A43676
15	39	50.0	444	2	microtubular asse
16	39	50.0	450	2	AC1938
17	39	50.0	469	2	G64361
18	39	50.0	503	2	G85432
19	39	50.0	655	2	S89836
20	39	50.0	655	2	S89836
21	38.5	49.4	567	2	S50708
22	38	48.7	68	2	UQ2005
23	38	48.7	190	2	A83102
24	38	48.7	234	2	A33170
25	38	48.7	257	2	F70166
26	38	48.7	300	2	T16778
27	38	48.7	504	1	S23558
28	38	48.7	533	2	T13607
29	38	48.7	553	2	G90593

30	38	48.7	557	2	AF0370	probable sulfatase
31	38	48.7	720	2	T15756	hypothetical prote
32	38	48.7	913	2	T18503	hypothetical prote
33	38	48.7	1051	2	T18351	limp protein - Myc
34	38	48.7	1365	2	T30822	limp protein - Myc
35	37	47.4	166	2	F70562	hypothetical prote
36	37	47.4	181	2	B26530	cytochrome-c oxida
37	37	47.4	184	2	AH3581	alkyl hydroperoxid
38	37	47.4	226	2	AC0176	probable exported
39	37	47.4	301	2	T18672	hypothetical prote
40	37	47.4	313	2	S30954	minor tail protein
41	37	47.4	322	2	F72800	minor tail subunit
42	37	47.4	352	2	I51687	neurogenic differe
43	37	47.4	359	2	S35157	Delta6 fatty acid
44	37	47.4	366	2	H63303	iron (III) ABC tra
45	37	47.4	398	2	S06324	dnab protein homol

ALIGNMENTS

RESULT 1  
PC1232  
copia polypeptide - fruit fly (Drosophila simulans) retrotransposon copia (fragments)  
C/Species: Drosophila simulans  
C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 10-Feb-1999  
C/Accession: PC1232  
R/Yoshiooka, K.; Kanda, H.; Takamatsu, N.; Togashi, S.; Kondo, S.; Miyake, T.; Sakaki, Y  
Gene 120, 191-196, 1992  
A/Title: Efficient amplification of Drosophila simulans copia directed by high-level re  
A/Reference number: PC1232; PMID:93013034; PMID:1383092  
A/Accession: PC1232  
A/Molecule type: DNA  
A/Residues: 1-313;314-787 <YOS>  
A/Cross-references: DDBJ:D10880  
C/Genetics:  
A/Genes: flyBase:copia  
A/Cross-references: FlyBase:FBgn0012867  
A/Mobile element: retrotransposon copia  
C/Superfamily: retrovirus-related polypeptide  
C/Keywords: polypeptide

Query Match 57.7%; Score 45; DB 2; Length 787;  
Best Local Similarity 69.2%; Pred. No. 10;  
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 IKRDHNDYSKNPM 14  
DB 183 IKRDHNDYSKNPM 195

RESULT 2  
OFFFCP  
copia polypeptide - fruit fly (Drosophila melanogaster) retrotransposon copia  
N/Contents: copia protein, 31k; copia protein, 48k; proteinase  
C/Species: Drosophila melanogaster  
C/Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 16-Jul-1999  
C/Accession: A03324; S03612; S14835  
R/Mount, S.M.; Rubin, G.M.  
Mol. Cell. Biol. 5, 1630-1638, 1985  
A/Title: Complete nucleotide sequence of the Drosophila transposable element copia: hon  
A/Reference number: A03324; PMID:85267679; PMID:2410772  
A/Accession: A03324  
A/Molecule type: DNA  
A/Residues: 1-1409 <MOU>  
A/Cross-references: GB:M1240; NID:G158615; PIDN:AA74497.1; PID:G950318  
Nucleic Acids Res. 17, 2134, 1989  
A/Title: The nucleotide sequence of Drosophila melanogaster copia-specific 2.1-kb mRNA.  
A/Reference number: S03612; PMID:89183629; PMID:2538806  
A/Accession: S03612  
A/Molecule type: mRNA  
A/Residues: 1-331,1375-1409 <ML>

A. Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauee, S.; Koetter, P.; Komnitsenos, G.; Krogan, S.; Kumano, M.; Kunita, K.; Lapidus, A.; Lathrop, S.; Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelli, L.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, A. Authors: Schleich, S.; Schroeter, R.; Scifione, F.; Sekiguchi, J.; Sekowska, A.; Serotoni, S.

Query Match	51.3%	Score 40;	DB 2;	Length 202;
Best Local Similarity	87.5%;	Pred. No. 18;		
Matches	7;	Conservative	0;	Mismatches 1;
				Indels 0;
				Gaps 0;



```
QY      4 RDHNDYSK 11
      181 RTDNDYSK 188

RESULT 7
E70184
ribose/galactose ABC transporter, permease protein (bdc-1) homolog - Lyme disease spiro
C/Species: Borrelia burgdorferi (Lyme disease spirochete)
C/Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 15-Jun-2001
C/Accession: E70184
R/Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Iachigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vingt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A/Authors: Smith, H.O.; Venter, J.C.
A/Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
A/Reference number: A70100; MUID:98065943; PMID:9403665
A/Accession: E70184
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-383 <KLE>
A/Cross-references: GB:AE001168; GB:AE000783; NID:G2688598; PIDN:AA67017.1; PID:G268860
C/Experimental source: strain B31
C/Superfamily: Archaeoglobus fulgidus ABC transporter AF0888

Query Match      51.3%; Score 40; DB 2; Length 383;
Best Local Similarity 70.0%; Pred. No. 35;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      2 IKRDNDYSK 11
      172 IKRDNDYSK 181

RESULT 8
T36881
probable transcription regulator - Streptomyces coelicolor
C/Species: Streptomyces coelicolor
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 07-Dec-1999
C/Accession: T36881
R/Murphy, L.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1999
A/Reference number: Z21617
A/Accession: T36881
A/Status: preliminary; translated from GB/EMBL/DD83
A/Molecule type: DNA
A/Residues: 1-548 <MUR>
A/Cross-references: EMBL:AL109848; PIDN:CA852842.1; GSPDB:GN00070; SCOEDB:SC151.18
A/Experimental source: strain A3(2)
C/Genetics:
A/Gene: SCOEDB:SC151.18

Query Match      51.3%; Score 40; DB 2; Length 548;
Best Local Similarity 63.6%; Pred. No. 51;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      4 RDHNDYSK 14
      475 RDHNDYSK 485

RESULT 9
F64662
trans-Golgi membrane protein p230 - Helicobacter pylori (strain 26695)
C/Species: Helicobacter pylori
C/Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C/Accession: F64662
R/Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodek, A.; McKen
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Wathey, L.
Nature 388, 539-547, 1997
A/Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.

A/Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A/Reference number: A64520; MUID:97394467; PMID:9252185
A/Accession: F64662
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-759 <TOM>
A/Cross-references: GB:AE000620; GB:AE000511; NID:G2314293; PIDN:AAD08188.1; PID:G23142
C/Genetics:
A/Start codon: GTG

Query Match      50.6%; Score 39.5; DB 2; Length 759;
Best Local Similarity 44.4%; Pred. No. 88;
Matches 8; Conservative 4; Mismatches 1; Indels 5; Gaps 1;

QY      2 IKRDNDYSK 14
      627 IKRDNDYSK 644

RESULT 10
F95159
prephenate dehydrogenase [imported] - Streptococcus pneumoniae (strain TIGR4)
C/Species: Streptococcus pneumoniae
C/Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 17-Feb-2003
C/Accession: F95159
R/Tettelein, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzaple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A/Authors: Loftus, B.O.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison
A/Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A/Reference number: A95000; MUID:21357209; PMID:11463916
A/Accession: F95159
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-367 <KUR>
A/Cross-references: GB:AE005672; PIDN:AAK75471.1; PID:G14972859; GSPDB:GN00164; TIGR:SP
A/Experimental source: strain TIGR4
C/Genetics:
A/Gene: SPI373
C/Superfamily: prephenate dehydrogenase, feedback inhibition-sensitive

Query Match      50.0%; Score 39; DB 2; Length 367;
Best Local Similarity 87.5%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 IKRDNDYSK 9
      21 IKRDNDYSK 28

RESULT 11
F98025
prephenate dehydrogenase (EC 1.3.1.12) [imported] - Streptococcus pneumoniae (strain R6
C/Species: Streptococcus pneumoniae
C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 17-Feb-2003
C/Accession: F98025
R/Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.;
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A/Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.
A/Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A/Reference number: A97872; MUID:21429245; PMID:11544234
A/Accession: F98025
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-367 <KUR>
A/Cross-references: GB:AE007317; PIDN:AAL00035.1; PID:G15458868; GSPDB:GN00174
C/Genetics:
A/Gene: tyra
C/Superfamily: prephenate dehydrogenase, feedback inhibition-sensitive
C/Keywords: oxidoreductase
```

Query Match 50.0%; Score 39; DB 2; Length 367;  
 Best Local Similarity 87.5%; Pred. No. 50;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 IKRDNDY 9  
 |||||  
 DB 21 IKRDHPDY 28

RESULT 12  
 D90528

hypothetical protein MYPV\_1320 [imported] - Mycoplasma pulmonis (strain UAB CT1P)  
 C/Species: Mycoplasma pulmonis  
 C/Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Aug-2001  
 C/Accession: D90528  
 R/Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallison, F.; Moszer, I.;  
 Nucleic Acids Res. 29, 2145-2153, 2001  
 A/Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm  
 A/Reference number: A95512; MUID:21267165; PMID:11353084  
 A/Accession: D90528  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-383 <KIR>  
 A/Cross-references: GB:A1445566; PID:g14089545; PIDN:CAC13305.1; GSPDB:GN00153  
 A/Experimental source: strain UAB CT1P  
 C/Genetics:  
 A/Gene: MYPV\_1320  
 A/Genetic code: SGC3

Query Match 50.0%; Score 39; DB 2; Length 383;  
 Best Local Similarity 50.0%; Pred. No. 53;  
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 IKRDNDYSKNP 13  
 : : : : :  
 DB 190 LKAHYDYDKNP 201

RESULT 13

hypothetical protein P9K20\_23 [imported] - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C/Accession: C96816  
 R/Rheologis, A.; Ecker, U.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Com, D.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 ansen, N.F.; Hughes, B.; Hultar, L.  
 Nature 408, 816-820, 2000  
 A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Yu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A/Reference number: A66141; MUID:21016719; PMID:11130712  
 A/Accession: C96816  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-412 <STO>  
 A/Cross-references: GB:AEO05173; NID:g3834320; PIDN:AAC83036.1; GSPDB:GN00141  
 C/Genetics:  
 A/Gene: P9K20\_23  
 A/Map position: 1

Query Match 50.0%; Score 39; DB 2; Length 412;  
 Best Local Similarity 53.8%; Pred. No. 57;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 SIKRDNDYSKNP 13  
 : : : : :  
 DB 316 SLHADFDDYSKTP 328

RESULT 14

A43676  
 p4 hepatitis-associated antigen - chimpanzee  
 C/Species: Pan troglodytes (chimpanzee)  
 C/Date: 03-Mar-1993 #sequence\_revision 03-Mar-1993 #text\_change 07-May-1999  
 C/Accession: A43676; A43677  
 R/Takahashi, K.; Kitamura, N.; Shibui, T.; Kamizono, M.; Matsui, R.; Yoshiyama, Y.; Maed  
 J. Gen. Virol. 71, 2005-2011, 1990  
 A/Title: Cloning, sequencing and expression in Escherichia coli of cDNA for a non-A, non  
 A/Reference number: A43676; MUID:91011346; PMID:2170570  
 A/Accession: A43676  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-444 <TR>  
 A/Cross-references: GB:D90034  
 R/Honda, Y.; Kondo, T.; Maeda, T.; Yoshiyama, Y.; Yamada, E.; Shimizu, Y.K.; Shikata, T.  
 J. Gen. Virol. 71, 1999-2004, 1990  
 A/Title: Isolation and purification of a non-A, non-B hepatitis-associated microtubular  
 A/Reference number: A43677; MUID:91011345; PMID:2170569  
 A/Accession: A43677  
 A/Status: preliminary  
 A/Molecule type: protein  
 A/Residues: 14-22;108-128;200-214;231-240 <HON>

Query Match 50.0%; Score 39; DB 2; Length 444;  
 Best Local Similarity 53.8%; Pred. No. 61;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 SIKRDNDYSKNP 13  
 |||||  
 DB 286 SIKLNHDDYDSP 298

RESULT 15

S48218  
 microtubular aggregate protein - human  
 C/Species: Homo sapiens (man)  
 C/Date: 15-Jul-1995 #sequence\_revision 19-Oct-1995 #text\_change 19-Oct-1995  
 C/Accession: S48218  
 R/Kitamura, A.; Takahashi, K.; Okajima, A.; Kitamura, N.  
 Eur. J. Biochem. 224, 877-883, 1994  
 A/Title: Induction of the human gene for p44, a hepatitis-C-associated microtubular aggr  
 A/Reference number: S48218; MUID:95010078; PMID:7925411  
 A/Accession: S48218  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-444 <KIT>

Query Match 50.0%; Score 39; DB 2; Length 444;  
 Best Local Similarity 53.8%; Pred. No. 61;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 SIKRDNDYSKNP 13  
 |||||  
 DB 286 SIKLNHDDYDSP 298

Search completed: March 22, 2004, 07:01:17  
 Job time : 3.91051 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - protein search, using sw model

Run on: March 22, 2004, 06:31:13 ; Search time 0.536965 Seconds

(Without alignments)  
1357.597 Million cell updates/sec

Title: US-09-662-293-1

Perfect score: 78

Sequence: 1 STRKDNYSKNPM 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

141681

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	57.7	1409	1	COP1_DROME
2	42	53.8	354	1	LSB5_YEAST
3	42	53.8	900	1	GCAB_BACSU
4	41	52.6	230	1	VATE_CITRL
5	41	52.6	230	1	VATE_CITRN
6	41	52.6	436	1	MURE_WISBR
7	39	50.0	444	1	P44_PANTR
8	39	50.0	469	1	SELB_METUA
9	39	50.0	1418	1	SMC4_YEAST
10	38.5	49.4	567	1	HXT9_YEAST
11	38	48.7	504	1	MPRA_SOLTU
12	37	47.4	181	1	COX2_PAPR
13	37	47.4	210	1	GTF1_BUPBU
14	37	47.4	290	1	OXAZ_STAEP
15	37	47.4	312	1	VG06_BPML5
16	37	47.4	321	1	VG06_BPMD2
17	37	47.4	352	1	NDP1_XENLA
18	37	47.4	359	1	LLCD_SYNY3
19	37	47.4	446	1	SOX8_HUMAN
20	37	47.4	451	1	GPID_CHLMU
21	37	47.4	451	1	GPID_CHLTR
22	37	47.4	457	1	RHO_FICFR
23	37	47.4	461	1	SOX10_CHICK
24	37	47.4	464	1	SOX8_MOUSE
25	37	47.4	466	1	SOX10_HUMAN
26	37	47.4	466	1	SOX10_MOUSE
27	37	47.4	466	1	SOX10_RAT
28	37	47.4	470	1	SOX8_CHICK
29	37	47.4	478	1	T23C_BACTB
30	37	47.4	494	1	SOX9_CHICK
31	37	47.4	497	1	MURE_BUCAP
32	37	47.4	509	1	SOX9_HUMAN
33	37	47.4	509	1	SOX9_PANTR

## ALIGNMENTS

34	37	47.4	509	1	SOX9_PIG	O18896 sus scrofa
35	37	47.4	550	1	GRP2_RAT	Q920W0 ratius norv
36	37	47.4	586	1	SYR_NITRU	Q92XCI nitrosomona
37	37	47.4	906	1	SECA_RICPR	Q92CXX rickettsia
38	37	47.4	1144	1	MZAI_HUMAN	Q16706 homo sapien
39	37	47.4	1165	1	GAPI_DROME	P48423 drosophila
40	37	47.4	1460	1	N159_YEAST	P40477 saccharomyc
41	37	47.4	5171	1	BPEA_HUMAN	O94833 homo sapien
42	36	46.2	317	1	CC28_CANTL	P43063 candida alb
43	36	46.2	317	1	OMPT_ECO57	P58603 escherichia
44	36	46.2	317	1	OMPT_ECOLI	P09169 escherichia
45	36	46.2	323	1	HEM2_ECOLI	P15002 escherichia

## RESULT 1

ID	COP1_DROME	STANDARD	PRT	1409 AA
AC	P04146; Q03728; Q24280; Q24555; Q24585; Q24586; Q24587;			
DT	01-NOV-1986 (Rel. 03, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Copia protein [contains: Copia VLP protein; Copia protease (EC 3.4.23.-)]			
GN	COP1A			
OS	Drosophila melanogaster (Fruit fly)			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORM LONG).			
RX	MEDLINE=85267679; PubMed=2410772;			
RA	Mount S.M., Rubin G.W.;			
RT	"Complete nucleotide sequence of the Drosophila transposable element			
RT	copia: homology between copia and retroviral proteins.";			
RL	Mol. Cell. Biol. 5:1630-1638(1985).			
RN	[2]			
RP	SEQUENCE FROM N.A., SEQUENCE OF 2-10, AND ALTERNATIVE SPLICING.			
RX	MEDLINE=85240569; PubMed=2409449;			
RA	Emori Y., Shiba T., Kanaya S., Inoue S., Yuki S., Saigo K.;			
RT	"The nucleotide sequences of copia and copia-related RNA in Drosophila			
RT	virus-like particles.";			
RL	Nature 315:773-776(1985).			
RN	[3]			
RP	SEQUENCE FROM N.A. (ISOFORM SHORT).			
RX	MEDLINE=89183629; PubMed=2538806;			
RA	Miller K., Rosenbaum J., Zdrzeczna V., Pogo A.O.;			
RT	"The nucleotide sequence of Drosophila melanogaster copia-specific			
RT	2.1-kb mRNA.";			
RL	Nucleic Acids Res. 17:2134-2134(1989).			
RN	[4]			
RP	SEQUENCE FROM N.A. (ISOFORM SHORT), AND MUTAGENESIS OF ASP-292.			
RC	TISSUE=Larva;			
RX	MEDLINE=9015130; PubMed=1689241;			
RA	Yoshioke K., Honma H., Zusni M., Kondo S., Togashi S., Miyake T.,			
RA	Shiba T.;			
RT	"Virus-like particle formation of Drosophila copia through			
RT	autocatalytic processing.";			
RL	EMBO J. 9:535-541(1990).			
CC	-1- ALTERNATIVE PRODUCTS:			
CC	Event=Alternative splicing; Named isoforms=2;			
CC	Name=Long;			
CC	Isoid=P04146-1; Sequence=Displayed;			
CC	Name=Short;			
CC	Isoid=P04146-2; Sequence=VSP 005226;			
CC	-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A11.			
CC	-1- SIMILARITY: Contains 1 CCHC-type zinc finger.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

-----  
 DR EMBL; X04456; CAA28054.2; -  
 DR EMBL; X02599; CAA26444.1; -  
 DR EMBL; X02599; CAA26445.1; -  
 DR EMBL; X02600; CAA26446.1; -  
 DR EMBL; X02600; CAA26447.1; -  
 DR EMBL; X13719; CAA31977.1; -  
 DR EMBL; X54147; CAA38086.1; -  
 DR PIR; A03324; OFFPCP.  
 DR MEROPS; All.001; -  
 DR PDBase; FBgn001347; cofilin/GIP.  
 DR InterPro; IPR001969; Aspartate\_As.  
 DR InterPro; IPR001584; Rve.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF00665; rve; 1.  
 DR Pfam; PF00999; ZF-CCHC; 1.  
 DR PRINTS; PR00399; C2HCZNFINGER.  
 DR SMART; SM00343; ZNF\_C2HC; 1.  
 DR PROSITE; PS00141; ASP\_PROTEASE; FALSE\_NEG.  
 DR PROSITE; PS00158; ZF\_CCHC; 1.  
 DR Transposable element Hydrolase; Aspartyl protease; ATP-binding; Polypeptide; Alternative splicing; Polymorphism; Zinc-finger.  
 KW CHAIN 1 270  
 FT CHAIN 1 270  
 FT ZN FING 230 247  
 FT ACT SITE 292 292  
 FT VASPLIC 392 1374  
 FT  
 FT VARIANT 1265 1288  
 FT MUTAGEN 1289 1409  
 FT CONFLICT 191 191  
 FT CONFLICT 300 300  
 FT CONFLICT 866 866  
 SQ SEQUENCE 1409 AA; 162817 MW; BE89440763A47691 CRC64;  
 Query Match 57.7%; Score 45; DB 1; Length 1409;  
 Best Local Similarity 69.2%; Pred. No. 6.3;  
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 2 IKRDNDYSKNPM 14  
 183 IKRDNDYSKNPM 195

RESULT 2  
 LSBS YEAST STANDARD; PRT; 354 AA.  
 AC P23679; P87004; Q8N1W9;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE LAS seventeen binding protein 5 (LAS17-binding protein 5).  
 GN LSBS OR YCL034W OR YCL34W OR YCL186.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 CC NCBI\_TaxID=4932;  
 CC  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC MEDLINE=91377317; PubMed=1897318;  
 CC Rad M.R., Luetzenkirchen K., Xu G., Kleinhaus U., Hollenberg C.P.;  
 CC "The complete sequence of a 11,953 bp fragment from Clg on chromosome  
 CC III encompasses four new open reading frames."  
 CC Yeast 7:533-538 (1991).  
 CC [2]  
 CC SEQUENCE FROM N.A.

RC STRAIN=5288c;  
 RX MEDLINE=92244356; PubMed=1574125;  
 RA Oliver S.G., van der Aart Q.J.M., Agostoni-Carbone M.L., Aigle M.,  
 RA Alberghina L., Alexandri D., Antoine G., Anwar R., Ballesta J.P.G.,  
 RA Bent P., Berben G., Bergantino E., Bileu N., Bolle P.-A.,  
 RA Bolotin-Fukuhara M., Brown A.J.P., Brown R., Buhler J.-M.,  
 RA Carignan G., Chanet R., Contreras R., Couzet M., Daignan-Fornier B.,  
 RA De Haan M., Defoor E., Delgado M.D., Demolder J., Dolta C., Dubois B.,  
 RA Dujon B., Duesterhoeft A., Erdmann D., Esteban M., Faure F.,  
 RA Fairhead C.A., Fave G., Feldmann H., Fiers W.,  
 RA Francinues-Gallard M.-C., Franco L., Frontali L., Fukuhara H.,  
 RA Fuller L.J., Gent M.E., Gigot D., Gilliguet V., Glansdorff N.,  
 RA Goffeau A., Grenson M., Grisanti P., Grivell L.A., Haesemann M.,  
 RA Hatat D., Hegemann U.H., Herbert C.J., Hilger F., Hohmann S.,  
 RA Hollenberg C.P., Huse K., Iborra F., Indge K.J., Iono K., Jackson P.,  
 RA Jacq C., Jaquet M., James C.M., Jauniaux J.-C., Jia Y., Jimenez A.,  
 RA Kleinhaus U., Kreisel P., Lafranchi G., Lewis C., van der Linden C.G.,  
 RA Lucchini G., Lutzenkirchen K., Maat C., Mannheim G., Manzano M.B.,  
 RA Martegani E., Mathieu A., Maurer C.T.C., McConnell D., McKee R.A.,  
 RA Messenguy F., Mewes H.-W., Molemans F., Montagne M.A., Navas L.,  
 RA Newlon C.S., Olson M.V., Paillet C., Panzeri L., Pearson B.M.,  
 RA Perea U., Phillipsen P., Pierard A., Planta R.J., Plevani P.,  
 RA Poetsch B., Pohl F.M., Purnelle B., Ramezani Rad M., Rasmussen S.W.,  
 RA Raynal A., Remacha M., Richeirich P., Roberts A.B., Rodriguez F.,  
 RA Sanz E., Schaef-gerstenschlaeger I., Scherens B., Schweitzer B.,  
 RA Shu Y., Skala J., Slonimski P.P., Sor F., Soustelle C.,  
 RA Spiegelberg R., Stata L.I., Steensma H.V., Steiner S., Thierry A.,  
 RA Thireos G., Triano L.N., Utrerasazu L.A., Valle G., Vetter I.,  
 RA van Vleet-Reedijk J.C., Volckaert G., Vreken P., Warrington J.R.,  
 RA von Wettstein D., Wicksteed B.L., Wilson C., Wurst H., Xu G.,  
 RA Zimmermann F.K., Sgouras J.G.,  
 RT "The complete DNA sequence of yeast chromosome III."  
 RL Nature 357:38-46(1992).  
 RN  
 RN REVISIONS.  
 RP Grossack R.;  
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP REVISIONS.  
 RA Valles G., Volckaerts G.;  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP PARTIAL SEQUENCE FROM N.A., FUNCTION, INTERACTION WITH SLA1, AND  
 RP SUBCELLULAR LOCATION.  
 RX MEDLINE=22287476; PubMed=12389763;  
 RA Dewar H., Warren D.T., Gardiner F.C., Gourlay C.G., Satish N.,  
 RA Richardson M.R., Andrews P.D., Ayecough K.R.;  
 RT "Novel proteins linking the actin cytoskeleton to the endocytic  
 RT machinery in Saccharomyces cerevisiae."  
 RL Mol. Biol. Cell 13:3646-3661(2002).  
 RN [6]  
 RP INTERACTION WITH LAS17.  
 RX MEDLINE=99443803; PubMed=10512884;  
 RA Madania A., Dimoulis P., Grava S., Kitamoto H., Schaner-Brobeck C.,  
 RA Souliard A., Moreau V., Winsor B.;  
 RT "The Saccharomyces cerevisiae homolog of human Wiskott-Aldrich  
 RT syndrome protein Las17 interacts with the Arp2/3 complex."  
 RL Mol. Biol. Cell 10:3521-3538(1999).  
 CC -1- FUNCTION: Essential for the organization of the actin  
 CC cytoskeleton, fluid phase endocytosis and vesicle trafficking,  
 CC together with YSC84.  
 CC -1- SUBUNIT: Interacts with SLA1 and LAS17.  
 CC -1- SUBCELLULAR LOCATION: Cell cortex. Localizes independently of F-  
 CC actin.  
 CC -1- SIMILARITY: TO S. POMBE SPBG31F10.07.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

```

CC -----
DR EMBL: X59720; CAC42957.1; -.
DR PIR: S74282; S74282.
DR GenBank: U00001; 138870; -.
DR SGD: S00005939; LSBS.
DR GO: GO:0005938; C:cell cortex; IDA.
DR GO: GO:0007015; P:actin filament organization; IPI.
DR GO: GO:0006897; P:cytoskeleton; IMP.
DR InterPro: IPR002014; VHS.
DR ProDom: PD003686; VHS; 1.
DR PROSITE: PS0179; VHS; FALSE_NEG.
DR Cytoskeleton: Endocytosis.
SQ SEQUENCE 354 AA; 39872 MW; 35772B575D40B82 CRC64;

Query Match
Best Local Similarity 77.8%; Score 42; DB 1; Length 354;
Pred. No. 4.8;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 IKRDNDY 10
DB 38 IKADNDY 46

RESULT 3
GGAB_BACSU STANDARD; PRT; 900 AA.
ID GGAB_BACSU
AC P46918;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Minor teichoic acid biosynthesis protein ggab.
GN GGAB OR BSU35680.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertolo M.G., Bessieres P., Bolicin A., Borcherdt S.,
RA Brouillet S., Brunsch C.V., Caldwell B., Capiano V., Carter N.M.,
RA Choi S.K., Codani U.J., Conerton J.F., Cummings N.J., Daniel R.A.,
RA Denholt F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Ertan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Gim S.Y., Glaeser P., Goffeau A., Golligly E.J., Grand G.,
RA Giuseppe G., Guy B.J., Haga K., Haeuch U., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.P., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasehara Y., Kjaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapilus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Melido R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Pato V., Poll T.M., Portetle D., Porwollik S., Prescott A.M.,
RA Pressen E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadie Y.,
RA Sao T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
RA Seifgich U., Sekowska K., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Taccoti E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpetra P., Tognoni A.,
RA Toato V., Uchiyama S., Vandenberg M., Vannier F., Vassarotti A.,
RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,

```

```

RA Yoshida K., Yoshikawa H.F., Zumsrein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RL Nature 390:249-256 (1997).
CC -1- FUNCTION: Involved in the biosynthesis of galactosamine-containing
CC minor teichoic acid, a non-essential cell wall polymer in
CC B. subtilis 168.
CC -1- PATHWAY: Poly(glycocyranose) N-acetylglucosamine 1-phosphate)
CC teichoic acid biosynthesis.
CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL: U13979; AA73513.1; -.
DR EMBL: Z99122; CAB15585.1; -.
DR PIR: E69631; E69631.
DR Subtilist; BG1192; ggab.
DR InterPro: IPR001173; Glyco_trans_2.
DR InterPro: IPR007554; Glyphos_transf.
DR Pfam: PF00535; Glycos_transf_2; 1.
DR Pfam: PF04464; Glyphos_transf_1.
DR KW Teichoic acid biosynthesis; Transferase; Glycosyltransferase;
KM Cell wall; Complete proteome.
SQ SEQUENCE 900 AA; 107154 MW; FA66495489C2C62F CRC64;

Query Match
Best Local Similarity 77.8%; Score 42; DB 1; Length 900;
Pred. No. 13;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SIKRDNDY 9
DB 399 NIKRDNDY 407

RESULT 4
VATE_CITR1 STANDARD; PRT; 230 AA.
ID VATE_CITR1
AC O95WE7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Vacuolar ATP synthase subunit E (EC 3.6.3.14) (V-ATPase E subunit)
DE (Vacuolar proton pump E subunit) (Clve-1).
GN VATE.
OS Citrus limon (Lemon).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Sapindales; Rutaceae; Citrus.
OX NCBI_TaxID=2708;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=fruit;
RA Reuveni M., Sadka A.;
RT "Isolation of vacuolar H+-ATPase subunit E cDNA from juice sacs of
RT Citrus limon."
RL (in) Plant Gene Register PGR00-005.
CC -1- FUNCTION: Subunit of the peripheral V1 complex of vacuolar ATPase
CC essential for assembly or catalytic function. V-ATPase is
CC responsible for acidifying a variety of intracellular compartments
CC in eukaryotic cells.
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +
CC H(+) (out).
CC -1- SUBUNIT: V-ATPase is a heteromultimeric enzyme composed of a
CC peripheral catalytic V1 complex (components A to H) attached to
CC an integral membrane V0 proton pore complex (components: a, c, c',
CC c'', and d).
CC -1- SIMILARITY: Belongs to the V-ATPase E subunit family.

```

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AF165939; AAD49706.1; -  
 DR InterPro; IPR002842; ATPsynth\_Esub.

DR Pfam; PF01991; V-ATP-synt\_E; 1.  
 KW Hydrolyase; ATP synthetase; Hydrogen ion transport.

SEQUENCE 230 AA; 26343 MW; 091360AE11DE068 CRC64;

Query Match 52.6%; Score 41; DB 1; Length 230;  
 Best Local Similarity 54.5%; Pred. NO. 4.5;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 SIKRDNDYSK 11  
 DB 102 NVSRDHSYK 112

# RESULT 5

V-ATP\_CITUN STANDARD; PRT; 230 AA.

AC Q9MBA6;  
 DT 16-OCT-2001 (Rel. 40; Created)  
 DT 16-OCT-2001 (Rel. 40; Last sequence update)  
 DT 28-EB-2003 (Rel. 41; Last annotation update)

DE Vacuolar ATP synthase subunit E (EC 3.6.3.14) (V-ATPase E subunit)  
 DE (Vacuolar proton pump E subunit).  
 DE VATE.

OS Citrus unshiu (Satsuma orange).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 CC eucosids II; Sapindales; Rutaceae; Citrus.

OK NCBI\_TaxID=55188;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RA Amentia T.; Komatsu A.; Yamamoto H.;

RT "Citrus unshiu mRNA for vacuolar H<sup>+</sup>-ATPase E subunit-1."  
 RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.

CC -1- FUNCTION: Subunit of the peripheral V1 complex of vacuolar ATPase  
 CC essential for assembly or catalytic function. V-ATPase is  
 CC responsible for acidifying a variety of intracellular compartments  
 CC in eukaryotic cells (By similarity).

CC -1- CATALYTIC ACTIVITY: ATP + H(2O) + H(+) (in) = ADP + phosphate +  
 CC H(+) (out).  
 CC -1- SUBUNIT: V-ATPase is a heteromultimeric enzyme composed of a  
 CC peripheral catalytic V1 complex (components A to H) attached to  
 CC an integral membrane V0 proton pore complex (components: a, c, c',  
 CC c'', and d) (By similarity).

CC -1- SIMILARITY: Belongs to the V-ATPase E subunit family.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AB037106; BA89661.1; -  
 DR InterPro; IPR002842; ATPsynth\_Esub.

DR Pfam; PF01991; V-ATP-synt\_E; 1.  
 KW Hydrolyase; ATP synthetase; Hydrogen ion transport.

SEQUENCE 230 AA; 26285 MW; AA1361E787C68BDE CRC64;

Query Match 52.6%; Score 41; DB 1; Length 230;  
 Best Local Similarity 54.5%; Pred. NO. 4.5;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 SIKRDNDYSK 11  
 DB 102 NVSRDHSYK 112

OY 1 SIKRDNDYSK 11  
 DB 102 NVSRDHSYK 112

# RESULT 6

MORE MIGR  
 ID MURR\_MIGR STANDARD; PRT; 496 AA.

AC Q8D221;  
 DT 10-OCT-2003 (Rel. 42; Created)  
 DT 10-OCT-2003 (Rel. 42; Last sequence update)  
 DT 10-OCT-2003 (Rel. 42; Last annotation update)

DE UDP-N-acetylmuramylalanyl-D-glutamate--2,6-diaminopimelate ligase  
 DE (EC 6.3.2.13) (UDP-N-acetylmuramyl-tripeptide synthetase) (meso-  
 DE diaminopimelate-adding enzyme) (UDP-MurNac-tripeptide synthetase).  
 GN MORE OR MIGR2110.

OS Wigglesworthia glosinidia brevipalpis.  
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC Enterobacteriaceae; Wigglesworthia.

CC NCBI\_TaxID=36870;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=22297718; PubMed=12219091;  
 RA Akman L.; Yamashita A.; Watanabe H.; Oshima K.; Shiba T.; Hattori M.,  
 RA Aksoy S.;

RT "Genome sequence of the endocellular obligate symbiont of tsetse  
 RT flies, Wigglesworthia glosinidia.";  
 RL Nat. Genet. 32:402-407(2002).

CC -1- FUNCTION: Cell wall formation. Diaminopimelic acid adding enzyme  
 CC (By similarity).  
 CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanyl-D-  
 CC glutamate + meso-2,6-diaminopimelate = ADP + phosphate + UDP-  
 CC N-acetylmuramoyl-L-alanyl-D-glutamate-glutamyl-meso-2,6-diamino-  
 CC heptanedioate.

CC -1- PATHWAY: Peptidoglycan biosynthesis.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC -1- SIMILARITY: Belongs to the murCDP family.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AB063521; BAC24357.1; -  
 DR HAMAP; MF\_00208; -; 1.

DR InterPro; IPR000713; Mur\_Ligase.  
 DR InterPro; IPR004101; Mur\_Ligase\_C.

DR InterPro; IPR005761; MURF.  
 DR Pfam; PF01225; Mur\_Ligase; 1.  
 DR Pfam; PF02875; Mur\_Ligase\_C; 1.

DR TIGRFAMs; TIGR01085; murF.1.  
 KW Peptidoglycan synthetase; Cell wall; Cell division; Ligase;  
 KW ATP-binding; Complete proteome.

FT NP BIND 118 124 ATP (POTENTIAL).  
 SQ SEQUENCE 496 AA; 57220 MW; AE624DA57ADDA1B CRC64;

Query Match 52.6%; Score 41; DB 1; Length 496;  
 Best Local Similarity 50.0%; Pred. NO. 10;  
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 SIKRDNDYSK 14  
 DB 208 NLSQDHDYK 221

RESULT 7  
 P44\_PANTR STANDARD; PRT; 444 AA.

ID P44\_PANTR  
 AC P27473;

DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, last sequence update)  
 DT 10-OCT-2003 (Rel. 42, last annotation update)  
 DE Non-A non-B hepatitis-associated microtubular aggregates protein  
 DE (Antigen p44).  
 OS Pan troglodytes (Chimpanzee).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.  
 ON NCBI\_TaxID=9598;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 14-22; 108-128 AND 231-240.  
 RC TISSUE=Liver;  
 RX MEDLINE=91011346; PubMed=2170570;  
 RA Takahashi K., Kitamura N., Shibui T., Kamizono M., Matsui R.,  
 RA Yoshiyama Y., Maeda T., Kondo J., Honda Y., Yamada E.,  
 RA Shimizu Y.K., Teranishi Y., Nakaniishi S.,  
 RT "Cloning, sequencing and expression in Escherichia coli of cDNA for a  
 RT non-A, non-B hepatitis-associated microtubular aggregates protein.";  
 RL J. Gen. Virol. 71:2005-2011(1990).  
 RN [2]  
 RP PARTIAL SEQUENCE.  
 RC TISSUE=Liver;  
 RX MEDLINE=91011345; PubMed=2170569;  
 RA Honda Y., Kondo J., Maeda T., Yoshiyama Y., Yamada E., Shimizu Y.K.,  
 RA Shikata T., Ono Y.,  
 RT "Isolation and purification of a non-A, non-B hepatitis-associated  
 RT microtubular aggregates protein.";  
 RL J. Gen. Virol. 71:1999-2004(1990).  
 CC -1- FUNCTION: This protein aggregates to form microtubular  
 CC structures.  
 CC -1- TISSUE SPECIFICITY: Hepatocytes.  
 CC -1- INDUCTION: During infection of non-A non-B hepatitis virus  
 CC or hepatitis delta virus.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D90034; BAA14082.1; -  
 DR Antigen; Microtubule.  
 SQ SEQUENCE 444 AA; 50471 MW; 799C6855018CFA71 CRC64;  
 Query Match 50.0%; Score 39; DB 1; Length 444;  
 Best Local Similarity 53.8%; Pred. No. 21;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 SIKRDNDYSKRP 13  
 DB 286 SIKLNHDYIDSP 298

RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weissbrock K.G., Merrick J.M., Glodok A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,  
 RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Weese C.R., Venter J.C.,  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 RT jannaschii";  
 RL Science 273:1058-1073 (1996).  
 RN [2]  
 RP FUNCTION.  
 RX MEDLINE=20318769; PubMed=10860743;  
 RA Rother M., Wiltberg R., Commans S., Boeck A.,  
 RT "Identification and characterisation of the selenocysteine-specific  
 RT translation factor SelB from the archaeon Methanococcus jannaschii";  
 RL J. Mol. Biol. 299:351-358(2000).  
 CC -1- FUNCTION: Translation factor necessary for the incorporation of  
 CC selenocysteine into proteins. It probably replaces EF-Tu for the  
 CC insertion of selenocysteine directed by the UGA codon. SelB binds  
 CC GTP and GDP.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: Belongs to the GTP-binding elongation factor family.  
 CC Self subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U67499; AAB98485.1; -  
 DR PIR; G64361; G64361.  
 DR HSP; P07157; IATP.  
 DR TIGR; M0495; -  
 DR InterPro; IPR000795; EF\_GTPbind.  
 DR InterPro; IPR004161; EFTU\_D2.  
 DR InterPro; IPR005225; Small GTP.  
 DR InterPro; IPR004535; TEF\_SelB.  
 DR InterPro; IPR009000; Translat\_factor.  
 DR Pfam; PRO0009; GTP\_EFTU; 1.  
 DR Pfam; PRO3144; GTP\_EFTU\_D2; 1.  
 DR PRINTS; PR00315; BLONGATNPT.  
 DR TIGRPM; TIGR00475; SelB; 1.  
 DR TIGRPM; TIGR00231; small GTP; 1.  
 DR PROSITE; PS00301; EFATOR\_GTP-binding; Complete proteome.  
 KW Protein biosynthesis; GTP-binding; Complete proteome.  
 FT NP\_BIND 16 23  
 FT NP\_BIND 69 73  
 FT NP\_BIND 123 126  
 FT NP\_BIND 123 126  
 SQ SEQUENCE 469 AA; 51849 MW; F3C1863CAAF0A92 CRC64;  
 Query Match 50.0%; Score 39; DB 1; Length 469;  
 Best Local Similarity 61.5%; Pred. No. 23;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 IRRDNDYSKRP 14  
 DB 182 IRRDINSYKMP 194

RESULT 9  
 SMCA YEAST  
 ID SMCA YEAST STANDARD; PRT; 1418 AA.  
 AC 012257;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, last sequence update)  
 DT 10-OCT-2003 (Rel. 42, last annotation update)  
 DE Structural maintenance of chromosome 4.  
 GN SMCA OR YLR086W OR L9449.5.  
 OS Saccharomyces cerevisiae (Baker's yeast).



CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 CC NCBI\_Taxid=4932;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN=SS286C / AB972;  
 CC MEDLINE=97313267; PubMed=9169871;  
 CC JOHNSON M., Hillier L., Riles L., Dubois E., Duesterhoeft A.,  
 CC Benes V., Brueckner M., Delius H., Goffeau A., Heblung U., Heumann K.,  
 CC Entian K.-D., Ploeth M., Goffeau A., Heblung U., Heumann K.,  
 CC Heuss-Neitzel D., Hilbert H., Hilger F., Kline K., Koetter P.,  
 CC Louis E.J., Messenguy F., Mewes H.-W., Miesga T., Moserl D.,  
 CC Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.W.,  
 CC Portetelle D., Purnelle B., Reckmann S., Rieger M., Rinke M., Rose M.,  
 CC Schaefer M., Scherens B., Scholler P., Schwager C., Schwarz S.,  
 CC Underwood A.P., Uristarazu L.A., Vandenbol M., Verhasselt P.,  
 CC Vierendeels F., Voet M., Volckaert G., Voss H., Wandt R., Wedler E.,  
 CC Wedler H., Zimmermann F.K., Zollner A., Hant U., Heisel J.D.,  
 CC "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.",  
 CC Nature 387:87-90(1997).  
 CC [2]  
 CC IDENTIFICATION IN A CONDENSIN COMPLEX WITH SMC2, BRN1, YCS4 AND  
 CC YCG1.  
 CC MEDLINE=20273907; PubMed=10811823;  
 CC Freeman L., Aragon-Alcalde L., Strunikov A.,  
 CC "The condensin complex governs chromosome condensation and mitotic  
 CC transmission of rDNA.",  
 CC J. Cell Biol. 149:811-824(2000).  
 CC -1- FUNCTION: Central component of the condensin complex, a complex  
 CC required for conversion of interphase chromatin into mitotic-like  
 CC condense chromosomes. The condensin complex probably introduces  
 CC positive supercoils into relaxed DNA in the presence of type I  
 CC topoisomerases and converts nicked DNA into positive knotted forms  
 CC in the presence of type II topoisomerases.  
 CC -1- SUBUNIT: Forms a heterodimer with SMC2. Component of the  
 CC condensin complex, which contains the SMC2 and SMC4 heterodimer,  
 CC and three non SMC subunits that probably regulate the complex:  
 CC BRN1, YCS4 and YCG1/YCS5.  
 CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic. In interphase  
 CC cells, the majority of the condensin complex is found in the  
 CC cytoplasm, while a minority of the complex is associated with  
 CC chromatin. A subpopulation of the complex however remains  
 CC associated with chromosome foci in interphase cells. During  
 CC mitosis, most of the condensin complex is associated with the  
 CC chromatin. At the onset of prophase, condensin associates with  
 CC chromosome arms and to chromosome condensation. Dissociation from  
 CC chromosomes is observed in late telophase.  
 CC -1- DOMAIN: The flexible hinge domain, which separates the large  
 CC intramolecular coiled coil regions, allows the heterodimerization  
 CC with SMC2, forming a V-shaped heterodimer (By similarity).  
 CC -1- SIMILARITY: Belongs to the SMC family. SMC4 subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL, 273258; CA97646.1; -  
 CC EMBL, 273259; CA97648.1; -  
 CC EMBL, U53880; AAB87590.1; -  
 CC PIR, S64918; S64918.  
 CC GenBank, 142148; -  
 CC SGD, S0004076; SMC4.  
 CC GO, GO:0000799; C:nuclear condensin complex; IPT.  
 CC GO, GO:0000706; P:mitotic chromosome condensation; IMP.  
 CC InterPro, IPR003439; ABC\_transporter.  
 CC InterPro, IPR003405; SMC\_C.  
 CC InterPro, IPR003395; SMC\_N.  
 CC Pfam, PF02483; SMC\_C; 1.  
 CC Pfam, PF02463; SMC\_N; 1.

KM DNA condensation; Mitosis; Cell cycle; ATP-binding; Coiled coil;  
 KW Nuclear protein.  
 FT NP\_BIND 185 192 ATP (POTENTIAL).  
 FT DOMAIN 345 673 COILED COIL (POTENTIAL).  
 FT DOMAIN 674 848 FLEXIBLE HINGE.  
 FT DOMAIN 849 1172 COILED COIL (POTENTIAL).  
 FT DOMAIN 1224 1263 COILED COIL (POTENTIAL).  
 FT DOMAIN 1323 1358 ALA/ASP-RICH (DA-BOX).  
 SQ SEQUENCE 1418 AA; 16218 MW; F0B6B72PEB5FD374 CRC64;  
 Query Match 50.0%; Score 39; DB 1; Length 1418;  
 Best Local Similarity 60.0%; Pred. No. 77;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 Qy 4 RDHDYSKNP 13  
 Db 101 RDHKSYSQSP 110  
 RESULT 10  
 HXT9 YEAST STANDARD; PRT; 567 AA.  
 ID HXT9 YEAST  
 AC P40885;  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hexose transporter HXT9.  
 GN HXT9 OR YJL219W OR J0222 OR HRC567.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 CC NCBI\_Taxid=4932;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN=SS286C; PubMed=7725802;  
 CC MEDLINE=95242842; PubMed=7725802;  
 CC Vandenbol M., Durand P., Bolle P.-A., Dion C., Portetelle D.,  
 CC Hilger F.,  
 CC "Sequence analysis of a 40.2 kb DNA fragment located near the left  
 CC telomere of yeast chromosome X.",  
 CC Yeast 10:11657-1662(1994).  
 CC -1- FUNCTION: PROBABLE GLUCOSE TRANSPORTER.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -1- SIMILARITY: Belongs to the sugar transporter family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL, 234098; CA9391.1; -  
 CC EMBL, 249494; CA93516.1; -  
 CC PIR, S50708; S50708.  
 CC GenBank, 141831; -  
 CC SGD, S0003755; HXT9.  
 CC GO, GO:0005353; F:fructose transporter activity; IGT.  
 CC GO, GO:0005354; F:galactose transporter activity; IGT.  
 CC GO, GO:0005355; F:glucose transporter activity; IGT.  
 CC GO, GO:0015578; F:mannose transporter activity; IGT.  
 CC InterPro, IPR007114; MFS.  
 CC InterPro, IPR005828; Sub\_transporter.  
 CC InterPro, IPR005829; Sug\_transporter.  
 CC InterPro, IPR003663; Sugar\_transpt.  
 CC Pfam, PF00083; sugar\_tr; 1.  
 CC PRINTS, PR00171; SUGRTNSFORT.  
 CC TIGRFAMs, TIGR00879; SP; 1.  
 CC PROSITE, PS00850; MFS; 1.  
 CC PROSITE, PS00216; SUGAR\_TRANSPORT\_1; 1.  
 CC PROSITE, PS00217; SUGAR\_TRANSPORT\_2; 1.



Repeat; Transmembrane; Sugar transport; Transport; Glycoprotein.

FT DOMAIN 1 56  
FT TRANSMEM 57 77  
FT TRANSMEM 78 112  
FT TRANSMEM 113 133  
FT TRANSMEM 134 139  
FT TRANSMEM 140 160  
FT TRANSMEM 161 170  
FT TRANSMEM 171 191  
FT TRANSMEM 192 197  
FT TRANSMEM 198 218  
FT TRANSMEM 219 232  
FT TRANSMEM 233 253  
FT TRANSMEM 254 336  
FT TRANSMEM 337 353  
FT TRANSMEM 354 359  
FT TRANSMEM 360 377  
FT TRANSMEM 378 384  
FT TRANSMEM 385 405  
FT TRANSMEM 406 429  
FT TRANSMEM 430 450  
FT TRANSMEM 451 467  
FT TRANSMEM 468 488  
FT TRANSMEM 489 499  
FT TRANSMEM 510 517  
FT TRANSMEM 511 567  
FT CARBOHYD 87 87  
FT CARBOHYD 227 227  
SQ SEQUENCE 567 AA; 62857 MW; 2CF4C8617D932701 CRC64;

Query Match 49.4%; Score 38.5; DB 1; Length 567;  
Best Local Similarity 57.1%; Pred. No. 34;  
Matches 8; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

Qy 1 SIKRDNDYSKNP 14  
Db 27 SVKTEHD-SKNSL 39

RESULT 11  
MPPA SOLTU STANDARD; PRT; 504 AA.  
AC P29677;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Mitochondrial processing peptidase alpha subunit, mitochondrial  
DE precursor (EC 3.4.24.64) (Alpha-MPP) (ubiquinol-cytochrome c reductase  
DE subunit II) (EC 1.10.2.2).  
GN MPP.  
OS Solanum tuberosum (Potato).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
OC Lamiales; Solanales; Solanaceae; Solanum.  
OX NCBI\_TaxID=4113;  
[1]  
SEQUENCE FROM N.A., AND SEQUENCE OF 270-280 AND 400-419.  
RX STRAIN=cv. Desiree; TISSUE=Green leaf;  
RC MEDLINE=92371428; Pubmed=1324169;  
RA Braun H.P., Emmertmann M., Knütt V., Schmitz U.K.;  
RT "The general mitochondrial processing peptidase from potato is an  
RT integral part of cytochrome c reductase of the respiratory chain.";  
RL EMBO J. 11:3219-3227(1992).  
CC -1- FUNCTION: Cleaves presequences (transit peptides) from  
CC mitochondrial protein precursors (By similarity).  
CC -1- FUNCTION: THIS IS A COMPONENT OF THE UBIQUINOL-CYTOCHROME C  
CC REDUCTASE COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH  
CC IS PART OF THE MITOCHONDRIAL RESPIRATORY CHAIN.  
CC -1- CATALYTIC ACTIVITY: Release of N-terminal transit peptides from  
CC precursor proteins imported into the mitochondrion, typically with  
CC Arg in position P2.  
CC -1- CATALYTIC ACTIVITY: OH(2) + 2 ferricytochrome c = O + 2

ferricytochrome c.  
CC -1- SUBUNIT: Heterodimer of alpha and beta subunits.  
CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane.  
CC -1- SIMILARITY: Belongs to peptidase family M16.  
CC -1- CAUTION: Does not seem to have a protease activity as it lack the  
CC zinc-binding site.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X66284; CAA6990.1; -.  
DR PIR; S23558; S23558.  
DR MEROPS; M16.971; -.  
DR InterPro; IPR001431; Peptidase\_M16.  
DR InterPro; IPR007863; Peptidase\_M16\_C.  
DR Pfam; PF00675; Peptidase\_M16\_1.  
DR Pfam; PF05193; Peptidase\_M16\_C\_1.  
DR PROSITE; PS00143; INSULINASE; 1.  
KM Hydroxylase; Metalloprotease; Mitochondrion; Transit peptide;  
KM Oxidoreductase; Electron transport; Respiratory chain.  
FT TRANSMEM 1 504  
FT TRANSMEM 504 AA; 54677 MW; FD3053688571D987 CRC64;  
SQ SEQUENCE 504 AA; 54677 MW; FD3053688571D987 CRC64;

Query Match 48.7%; Score 38; DB 1; Length 504;  
Best Local Similarity 50.0%; Pred. No. 37;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 2 IKRDNNDYSKNP 13  
Db 195 VKAEISEYSKNP 206

RESULT 12  
COX2\_PAPPR STANDARD; PRT; 181 AA.  
AC P08748;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Cytochrome c oxidase polypeptide II (EC 1.9.3.1).  
GN COII.  
OS Parametium primaurelia.  
OC Mitochondrion.  
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Periculiata;  
OC Parametium.  
OX NCBI\_TaxID=5886;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87131992; Pubmed=3032745;  
RA Mahalingam R., Seilhamer J.U., Pritchard A.E., Cummings D.U.;  
RT "Identification of Parametium mitochondrial proteins using antibodies  
RT raised against fused mitochondrial gene products.";  
RL Gene 49:129-138(1986).  
CC -1- FUNCTION: Cytochrome c oxidase is the component of the respiratory  
CC chain that catalyzes the reduction of oxygen to water. Subunits 1-  
CC 3 form the functional core of the enzyme complex. Subunit 2  
CC transfers the electrons from cytochrome c via its binuclear copper  
CC A center to the binuclear center of the catalytic subunit 1.  
CC -1- CATALYTIC ACTIVITY: 4 ferricytochrome c + O(2) = 4 ferricytochrome  
CC c + 2 H(2)O.  
CC -1- COFACTOR: Copper A.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial  
CC inner membrane.  
CC -1- SIMILARITY: Belongs to the cytochrome c oxidase subunit 2 family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL/ M15219; AAA32007.2; ALT\_SEQ.

DR PIR; B26530; B26530.  
 DR InterPro: IPR001505; Copper\_CuA.  
 DR InterPro: IPR008972; Cupredoxin.  
 DR InterPro: IPR002429; Cyt\_cox\_2.  
 DR Pfam; PF00116; COX2; 1.  
 DR PRINTS; PR01166; CYCOXIDASEII.  
 DR PRODOM; PD000131; Copper\_CuA; 1.  
 DR PROSITE; PS00078; COX2; FALSE NEG.  
 KW Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;  
 KW Electron transport; Respiratory chain.  
 FT TRANSMEM 80 100  
 FT METAL 126 126  
 FT METAL 130 130  
 FT METAL 134 134  
 FT METAL 134 134  
 SQ SEQUENCE 181 AA; 21232 MW; 910481DF1A455A76 CRC64;

Query Match 47.4%; Score 37; DB 1; Length 181;  
 Best Local Similarity 63.6%; Pred. No. 18;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 2 IKRDNDYSKN 12  
 Db 5 VLDKNDLSKN 15

## RESULT 13

GRP1\_BUPBU STANDARD; PRT; 210 AA.  
 ID GRP1\_BUPBU  
 AC P81942;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Glutathione S-transferase P 1 (EC 2.5.1.18) (BBGSP1-1) (GST class-  
 DE pi).  
 OS Bufo bufo (European toad).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Hylidae; Bufonidae; Bufo.  
 OX NCBI\_TaxID=8384;  
 RN (1)  
 RP SEQUENCE.  
 RP TISSUE=Embryo;  
 RC MEDLINE=97218132; PubMed=9065793;  
 RA Sacchetti P., Petruzzelli R., Melino S., Buccirelli T., Pennelli A.,  
 RA Amicarelli F., Miranda M., Di Ilio C.;  
 RT "Amphibian embryo glutathione transferase: amino acid sequence and  
 RT structural properties";  
 RL Biochem J 332:679-680(1997).  
 CC -1- FUNCTION: Conjugation of reduced glutathione to a wide number  
 CC of exogenous and endogenous hydrophobic electrophiles.  
 CC -1- CATALYTIC ACTIVITY: RX + glutathione = HX + R-S-glutathione.  
 CC -1- SUBUNIT: Homodimer.  
 CC -1- SIMILARITY: Belongs to the GST superfamily. Pf family.

CC HSSP; PI9157; 1BAY.  
 DR InterPro; IPR004046; GST\_Cterm.  
 DR InterPro; IPR004045; GST\_Nterm.  
 DR InterPro; IPR003082; GST\_pi.  
 DR Pfam; PF00043; GST\_C\_1.  
 DR Pfam; PF02798; GST\_N\_1.  
 DR PRINTS; PR01266; GSTXSPRASEP.  
 KW Transferase; Multigene family.  
 SQ SEQUENCE 210 AA; 23820 MW; AC835C0390F3A71D CRC64;

Query Match 47.4%; Score 37; DB 1; Length 210;  
 Best Local Similarity 50.0%; Pred. No. 21;  
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Oy 2 IKRDNDYSKN 13  
 Db 72 IARNHDLGKN 83

## RESULT 14

OXK2\_STAEP STANDARD; PRT; 290 AA.  
 ID OXK2\_STAEP  
 AC O8CMK8;  
 DT 15-MAR-2004 (Rel. 43, Created)  
 DT 15-MAR-2004 (Rel. 43, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Membrane protein oxa2 2 precursor.  
 GN OXA2 OR S81689.  
 OS Staphylococcus epidermidis.  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxID=1282;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 12228;  
 RX MEDLINE=22832016; PubMed=12950922;  
 RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,  
 RA Qiu Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,  
 RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;  
 RT "Genome-based analysis of virulence genes in a non-biofilm-forming  
 RT Staphylococcus epidermidis strain (ATCC 12228).";  
 RL Mol. Microbiol. 49:1577-1593(2003).

CC -1- FUNCTION: Required for the insertion of integral membrane proteins  
 CC into the membrane. May also be involved in protein secretion  
 CC processes (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
 CC -1- SIMILARITY: Belongs to the OXA1/oxa family. Subfamily 2.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; AE016749; AAC05288.1; -.  
 DR HAMAP; MF 01811; -1.  
 DR InterPro; IPR001708; 60kDa\_innermemb.  
 DR InterPro; IPR000437; Prok\_lipoprot\_S.  
 DR Pfam; PF02096; 60KD\_LMP; 1.  
 DR PRINTS; PR00701; 60KDINNERMP.  
 DR PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.  
 KW Transmembrane; Lipoprotein; Signal; Palmitate; Complete proteome.

FT SIGNAL 1 19  
 FT CHAIN 20 290  
 FT TRANSMEM 57 79  
 FT TRANSMEM 135 154  
 FT TRANSMEM 178 195  
 FT TRANSMEM 207 224  
 FT TRANSMEM 229 251  
 FT TRANSMEM 20 20  
 FT LIPID 20 20  
 FT LIPID 20 20  
 SQ SEQUENCE 290 AA; 33747 MW; 913063238F04D09E CRC64;

Query Match 47.4%; Score 37; DB 1; Length 290;  
 Best Local Similarity 66.7%; Pred. No. 30;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 3 IKRDNDYSK 11  
 Db 267 ERENHPSK 275

RESULT 15  
 VG06\_BPMU5 STANDARD; PRT; 312 AA.  
 ID VG06\_BPMU5

AC Q05278; 01-FRB-1994 (Rel. 28, Created)  
 DT 01-FRB-1994 (Rel. 28, Last sequence update)  
 DT 01-FRB-1994 (Rel. 28, Last annotation update)  
 DE Minor tail protein GP6.  
 GN 6.  
 OS Mycobacteriophage L5.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;  
 OC L5-like viruses.  
 OX NCBI\_TaxID=31757;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 1-11.  
 RX MEDLINE=93211282; Pubmed=8459766;  
 RA Hatfull G.F., Sarkis G.J.;  
 RT "DNA sequence, structure and gene expression of mycobacteriophage L5:  
 RL Mol. Microbiol. 7:395-405(1993).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; 218946; CAA79385.1; -.  
 DR PIR; S30954; S30954.  
 FT INIT MET 0  
 SQ SEQUENCE 312 AA; 33890 MW; 0D392560248349B3 CRC64;

Query Match 47.4%; Score 37; DB 1; Length 312;  
 Best Local Similarity 66.7%; Pred. No. 33;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 DHNDYSKMP 13  
 | : | | | : | |  
 Db 72 DYNDVSENP 80

Search completed: March 22, 2004, 06:52:53  
 Job time : 2.53696 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:39:53 ; Search time 2.81712 Seconds  
(without alignments)  
1568.003 Million cell updates/sec

Title: us-09-662-293-1  
Perfect score: 78  
Sequence: 1 SIKRDNDYKNPM 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues  
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phage:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_virus:\*
  - 16: sp\_bacteriaph:\*
  - 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	78	100.0	555	5	Q9U6R7 dermatophag
2	46	59.0	569	5	Q81L54 plasmodium
3	45	57.7	239	5	Q9U6V8
4	45	57.7	445	4	Q9H8K7
5	45	57.7	1250	5	Q813V2
6	45	57.7	1409	5	Q08461
7	44	56.4	227	2	Q9S489
8	44	56.4	227	16	Q7V2B7
9	44	56.4	973	5	Q81KX6
10	42	53.8	576	16	Q7UGM6
11	42	53.8	601	11	Q91YV4
12	42	53.8	777	11	Q8C9F1
13	42	53.8	840	3	Q9K057
14	42	53.8	378	5	Q9N4X7
15	41	52.6	441	16	Q8DAK4
16	41	52.6	441	16	Q8DAK4

17	41	52.6	748	5	Q81KY2	plasmodium
18	41	52.6	899	5	Q81H57	plasmodium
19	41	52.6	906	2	Q7X081	ricketsia
20	41	52.6	906	16	Q92H92	dictyostella
21	41	52.6	920	5	Q86K51	dictyostella
22	41	52.6	1645	10	Q7X642	oryza sativ
23	41	52.6	1923	5	Q81BH8	plasmodium
24	40	51.3	1802	17	Q8TSH2	methanosarc
25	40	51.3	202	16	Q9ZEP5	streptomyce
26	40	51.3	228	5	Q25361	borrelia bu
27	40	51.3	383	16	Q51621	borrelia bu
28	40	51.3	445	4	Q8N597	homio sapien
29	40	51.3	470	16	Q7WQW7	borrelia
30	40	51.3	470	16	Q7WQW7	borrelia
31	40	51.3	470	16	Q7WQW7	borrelia
32	40	51.3	505	8	Q7YKX5	uricularia
33	40	51.3	548	16	Q9S222	streptomyce
34	40	51.3	837	5	Q95P23	aplysia cal
35	40	51.3	1621	5	Q81IL7	plasmodium
36	39.5	50.6	296	10	Q7XES0	oryza sativ
37	39.5	50.6	548	5	Q813H5	plasmodium
38	39.5	50.6	575	5	Q816Q2	anopheles s
39	39.5	50.6	759	16	Q25761	helicobacte
40	39	50.0	112	16	Q8EEK2	shewanella
41	39	50.0	156	16	Q88C16	pseudomonas
42	39	50.0	206	12	Q7TSN9	cryptophleb
43	39	50.0	273	2	Q54270	staphylococ
44	39	50.0	284	3	Q9C125	pichia etch
45	39	50.0	342	12	Q91F03	cydia pomon

ALIGNMENTS

RESULT 1

Q9U6R7 PRELIMINARY; PRT; 555 AA.

AC Q9U6R7; 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE 98kDa HDM allergen.

OS Dermatophagoides farinae (House-dust mite).

OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;

OC Acariformes; Sarcophormes; Astigmata; Psoroptida; Analgoidea;

OC Pyroglyphidae; Dermatophagoides.

OX NCBI\_TaxID=6954;

RN [1]

RP SEQUENCE FROM N.A.

RA Weber E.R., Hunter S., Steadman K., McCall C.;

RT "Cloning and Characterization of a 98 kDa Allergen from Dermatophagoides farinae."

RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF178772; AAD52672.1; -

DR GO; GO:0005576; P:chitin metabolism; IEA.

DR GO; GO:0005798; P:chitin binding; IEA.

DR GO; GO:0005975; P:carbohydrate metabolism; IEA.

DR GO; GO:0006030; P:chitin metabolism; IEA.

DR InterPro; IPR002557; Chitin bind per.

DR InterPro; IPR001223; Glyco\_hydro\_18.

DR InterPro; IPR001579; Glyco\_hydro\_18A.

DR Pfam; PF00704; Glyco\_hydro\_18; 1.

DR ProDom; PD000471; Glyco\_hydro\_18; 1.

DR SMART; SM00636; Chitinase; 1.

DR PROSITE; PS01095; CHITINASE\_18; 1.

KW Glycosylase; Hydrolase.

SQ SEQUENCE 555 AA; 63238 MW; 0E4564A1A59830B CRC64;

Query Match 100.0%; Score 78; DB 5; Length 555;

Best Local Similarity 100.0%; Pred. No. 5.6e-05;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SKRDNDYKSNP 14  
 Db 20 SKRDNDYKSNP 33

## RESULT 2

Q81L54 PRELIMINARY; PRT; 569 AA.  
 AC Q81L54;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Hypothetical protein, conserved.  
 GN PF14\_0395.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=22255705; PubMed=12368864;  
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W., James K., Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K., Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S., Chan M.-S., Nene V., Shallow S.J., Sub B., Peterson J., Angiuoli S., Pereira M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B., Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A., McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C., Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W., Fraser C.M., Barrell B.;  
 RA "Genome sequence of the human malaria parasite Plasmodium falciparum.";  
 RT Nature 419:498-511(2002).  
 RL EMBL; AF014822; AAC37008.1;  
 DR GO; GO:0003824; F: catalytic activity; IEA.  
 DR InterPro; IPR000379; Ser\_estrs.  
 KW Hypothetical protein.  
 SQ SEQUENCE 569 AA; 67622 MW; FB86C553DBD1F4C7 CRC64;  
 QY 3 KRDNDYKSNP 13  
 Db 327 KRDNDYKSNP 337

## RESULT 3

Q9U6V8 PRELIMINARY; PRT; 239 AA.  
 AC Q9U6V8;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Polypeptide (Fragment).  
 GN COP1A\GIP.  
 OS Drosophila willistoni ('Fruit fly').  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidae; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7260;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Royal Palm Park; TRANSPOSON=copia;  
 RX MEDLINE=20006284; PubMed=10535972;  
 RA Jordan I.K., Matyunina L.V., McDonald J.F.;  
 RT "Evidence for the recent horizontal transfer of long terminal repeat retrotransposon.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:12621-12625(1999).  
 DR EMBL; AF175766; AAF06364.1; -;  
 DR Flybase; FBgn0044137; Dm11\copia\GIP.

FT NON TER 239 239  
 SQ SEQUENCE 239 AA; 27400 MW; FA481F22AFB91C50 CRC64;

Query Match 57.7%; Score 45; DB 5; Length 239;  
 Best Local Similarity 69.2%; Pred. No. 11;  
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 IKRDNDYKSNP 14  
 Db 183 IKRDNDYKSNP 195

## RESULT 4

Q9H8K7 PRELIMINARY; PRT; 445 AA.  
 AC Q9H8K7;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Hypothetical protein FLJ13490.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Ilogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y., Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Nagasuna M., Hosoki T., Kaku Y., Kodaira H., Kondo H., Sugawara M., Takahashi M., Chida Y., Ishida S., Murakawa K., Ono Y., Takiguchi S., Watanabe S., Kimura K., Murakami K., Ishi S., Kawai Y., Saito K., Yamamoto J., Wakatsuki A., Nakamura Y., Nagahara K., Masuno Y., Nishimura K., Iwayanagi T.;  
 RA "MDO human cDNA sequencing project.";  
 RT Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; AK023552; BAB14609.1; -;  
 DR GO; GO:0016020; C: membrane; IEA.  
 DR GO; GO:0005524; F: ATP binding; IEA.  
 DR GO; GO:0004009; F: ATP-binding cassette (ABC) transporter acti. .; IEA.  
 DR GO; GO:0006810; P: transport; IEA.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 445 AA; 49250 MW; D08BE83124EE5189 CRC64;

Query Match 57.7%; Score 45; DB 4; Length 445;  
 Best Local Similarity 58.3%; Pred. No. 21;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 IKRDNDYKSNP 13  
 Db 61 IKRDNDYKSNP 72

## RESULT 5

Q813V2 PRELIMINARY; PRT; 1250 AA.  
 AC Q813V2;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN PFE0780W.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Devlin K., Baker S., Churcher C., Quail M., Barrell B.;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=2255708; PubMed=12368867;  
 RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,  
 RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,  
 RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,  
 RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Deggett J.,  
 RA Felwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,  
 RA Harper D., Hauser H., Hornsby T., Holtroyd S., Horrocks P.,  
 RA Humphray S., Jagsels K., James K.D., Johnson D., Kenchou A.,  
 RA Knights A., Konfortov B., Kyes S., Larkie N., Lawson D., Lennard N.,  
 RA Line A., Maddison M., Mclean J., Mooney P., Moute S., Murphy L.,  
 RA Oliver K., Ormond D., Price C., Quail M.A., Rabinovitch E.,  
 RA Rajadream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,  
 RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,  
 RA Sulston J.E., Craig A., Newbold C., Barrell B.G;  
 RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."  
 RL Nature 419:527-531(2002).  
 DR EMBL: AL929352; CAD51520.1; -.  
 KW Hypothetical protein.  
 SO SEQUENCE 1250 AA; 146736 MW; 0DA30597175AC21A CRC64;

Query Match 57.7%; Score 45; DB 5; Length 1250;  
 Best Local Similarity 77.8%; Pred. No. 63;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 RDHNDYSKN 12  
 Db 1216 KDHNDYSKN 1224

RESULT 6  
 Q08461 PRELIMINARY; PRT; 1409 AA.  
 AC Q08461;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Copia.  
 GN COPIA\GIP.  
 OS Drosophila simulans (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7240;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93013034; PubMed=1381092;  
 RA Yoshioke K., Kanda H., Takamatsu N., Togaehi S., Kondo S., Miyake T.,  
 RA Sakaki Y., Shiba T.;  
 RT "Efficient amplification of Drosophila simulans copia directed by  
 RT high-level reverse transcriptase activity associated with copia virus-  
 RT like particle."  
 RL Gene 120:191-196(1992).  
 DR EMBL: D10880; AAA01703.1; -.  
 DR MEROPS; A11.001; -.  
 DR FLYBASE: FBgn0044190; Dsim\copia\GIP.  
 DR GO: GO:0003677; F:DNA binding; IEA.  
 DR GO: GO:0006310; P:DNA recombination; IEA.  
 DR InterPro; IPR001584; RVE.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF00665; rve; 1.  
 DR Pfam; PF00098; zf-CCHC; 1.  
 DR PRINTS; PRO0939; CCHCNFNGER.  
 DR SMART; SM00343; ZNF\_CCHC; 1.  
 DR PROSITE; PS50158; ZF\_CCHC; 1.  
 FT CHAIN 1 270 MAJOR COPIA VIRUS-LIKE PARTICLES PROTEIN.  
 SO SEQUENCE 1409 AA; 162899 MW; 403A9FF66A21903 CRC64;

QY 2 IKRDHNDYSKNPM 14  
 Db 183 IKDHNDYSKKVM 195

RESULT 7  
 Q95489 PRELIMINARY; PRT; 227 AA.  
 AC Q95489;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Prochlorococcus sp. (strain PCC 9511).  
 OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococaceae;  
 CC Prochlorococcus.  
 OX NCBI\_TaxID=100365;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PCC 9511;  
 RX MEDLINE=21142517; PubMed=11208789;  
 RA Holtzendorf J., Pattersky F., Jacquet S., Bryant F., Marie D.,  
 RA Garzatek L., Mary I., Vault D., Hess W.R.;  
 RT "Diels Expression of Cell Cycle-Related Genes in Synchronized Cultures  
 RT of Prochlorococcus sp. Strain PCC 9511."  
 RL J. Bacteriol. 183:915-920(2001).  
 DR EMBL: AF158628; AAD45693.1; -.  
 DR InterPro; IPR001173; Glyco\_trans\_2.  
 DR Pfam; PF00535; Glycos\_transf\_2; 1.  
 KW Hypothetical protein.  
 SO SEQUENCE 227 AA; 27203 MW; 5E1EB4F18559F019 CRC64;

Query Match 56.4%; Score 44; DB 2; Length 227;  
 Best Local Similarity 75.0%; Pred. No. 15;  
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SIKRDHNDYSKN 12  
 Db 216 SIKSLHNDYSKN 227

RESULT 8  
 Q7V2B7 PRELIMINARY; PRT; 227 AA.  
 AC Q7V2B7;  
 DT 01-OCT-2003 (TREMBLrel. 25, Created)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN PMM0563.  
 OS Prochlorococcus marinus subsp. pastoris (strain CCMP 1378 / MED4).  
 OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococaceae;  
 CC Prochlorococcus.  
 OX NCBI\_TaxID=59919;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22825698; PubMed=12917642;  
 RA Rosp G., Larimer F.W., Lamerdin J., Malfatti S., Chain P.,  
 RA Ahlgren N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,  
 RA Johnson Z.I., Land M., Lindell D., Post A.R., Regala W., Shah M.,  
 RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,  
 RA Webb E.A., Zinser E.R., Chisholm S.W.;  
 RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic  
 RT niche differentiation."  
 RL Nature 424:1042-1047(2003).  
 DR EMBL: BX572091; CAE19022.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 227 AA; 27204 MW; 5E1EB4F18559F019 CRC64;

Query Match 56.4%; Score 44; DB 16; Length 227;  
 Best Local Similarity 75.0%; Pred. No. 15;  
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 SIKRDNDYSKN 12  
 DB 216 SIKRDNDYSKN 227

## RESULT 9

Q8IKN6 PRELIMINARY; PRT; 973 AA.  
 AC O8IKN6;  
 DT 01-MAR-2003 (TRENBLrel. 23, Created)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN Pf14 0568.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=22255705; PubMed=12368864;  
 RA Gardner W.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,  
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
 RA Chan M.-S., Nene V., Shallow S.J., Sub B., Peterson J., Angiuoli S.,  
 RA Petrea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,  
 RA Martin D.M.A., Fairlamb A.H., Fraunholz M.V., Roos D.S., Ralph S.A.,  
 RA McRadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
 RA Fraser C.M., Barrell B.;  
 RT "Genome sequence of the human malaria parasite Plasmodium  
 RT falciparum."  
 RL Nature 419:498-511(2002).  
 DR EMBL; AE014825; AAN37181.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 973 AA; 118769 MW; EEB74D59493F443 CRC64;

Query Match 56.4%; Score 44; DB 5; Length 973;  
 Best Local Similarity 58.3%; Pred. No. 72;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 SIKRDNDYSKN 12  
 DB 300 NIKRDNDYSKN 311

## RESULT 10

Q7UGM6 PRELIMINARY; PRT; 576 AA.  
 AC Q7UGM6;  
 DT 01-OCT-2003 (TRENBLrel. 25, Created)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Probable flagellar P-ring protein.  
 GN FlgI OR RB5126.  
 OS Rhodospirillum rubrum.  
 OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;  
 OC Planctomycetaceae; Pirellula.  
 OX NCBI\_TaxID=117;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1;  
 RX MEDLINE=22735913; PubMed=12835416;  
 RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,  
 RA Ludwig W., Gade D., Beck A., Borzym K., Heilmann K., Rabus R.,  
 RA Schlesner H., Aumann R., Reinhardt R.;  
 RT "Complete genome sequence of the marine planctomycete Pirellula sp.  
 RT strain 1."  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).  
 DR EMBL; BX294141; CAD78303.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 576 AA; 62785 MW; 41537F159CC27E0B CRC64;

Query Match 53.8%; Score 42; DB 16; Length 576;  
 Best Local Similarity 72.7%; Pred. No. 91;  
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 KRQNDYSKNP 13  
 DB 32 KRQNDYSKNP 42

## RESULT 11

Q91YV4 PRELIMINARY; PRT; 601 AA.  
 AC Q91YV4;  
 DT 01-DEC-2001 (TRENBLrel. 19, Created)  
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 GN 4921524P20R1K.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; BC013814; AA013814.1; -.  
 DR MGD; MGI:1915509; 4921524P20R1K.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF00400; WD40; 2.  
 DR SMART; SM00320; WD40; 3.  
 DR PROSITE; PSS0082; WD\_REPEATS\_2; 1.  
 DR PROSITE; PSS0284; WD\_REPEAT\_REGION; 1.  
 KW Hypothetical protein; Repeat; WD repeat.  
 SQ SEQUENCE 601 AA; 68243 MW; F0470922F18A3C9B CRC64;

Query Match 53.8%; Score 42; DB 11; Length 601;  
 Best Local Similarity 63.6%; Pred. No. 95;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 KRQNDYSKNP 13  
 DB 375 KRQNDYSKNP 385

## RESULT 12

Q8C9F1 PRELIMINARY; PRT; 751 AA.  
 AC Q8C9F1;  
 DT 01-MAR-2003 (TRENBLrel. 23, Created)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
 DE Hypothetical Ttp-Aep (Fragment).  
 GN 4921524P20R1K.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Thymus;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The PANTOM Consortium.  
 RT "Analysis of the mouse transcription Research Group Phase I & II Team;  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK042231; BAC31200.1; -.  
 DR MGD; MGI:1915509; 4921524P20R1K.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF00400; WD40; 4.

```

DR PROSITE: PSS0082; WD_REPEATS_2; 2.
DR PROSITE: PSS0294; WD_REPEATS_REGION; 1.
KW Hypothetical protein.
FT NON_TER 751
SQ SEQUENCE 751 AA; 84910 MW; B78CABDACE7B6157 CRC64;

Query Match
Best Local Similarity 53.8%; Score 42; DB 11; Length 751;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 KRDNHYSKNP 13
DB 551 ERDASEYSKNP 561

RESULT 13
Q8K057 PRELIMINARY; PRT; 777 AA.
AC Q8K057;
DT 01-OCT-2002 (TRENBLREL. 22, Created)
DT 01-OCT-2002 (TRENBLREL. 22, Last sequence update)
DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
DE Similar to KIAA1374 protein.
GN 4921524P20R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC Straussberg R.;
RA Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL: BC034101; AAH34101.1; -
DR MGD; MGI:1915509; 4921524P20R1K.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 4.
DR PROSITE: PSS0082; WD_REPEATS_2; 2.
DR PROSITE: PSS0294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 777 AA; 87810 MW; 2B01621A12B60C4 CRC64;

Query Match
Best Local Similarity 53.8%; Score 42; DB 11; Length 777;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 KRDNHYSKNP 13
DB 551 ERDASEYSKNP 561

RESULT 14
Q9URY8 PRELIMINARY; PRT; 840 AA.
AC Q9URY8;
DT 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
DT 01-JUN-2003 (TRENBLREL. 24, Last annotation update)
DE Probable sulfate permease.
GN SPAC869.05C.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972h-1.
RA Hunt C., Aves S., McDougall R.C., Rajandream M.A., Barrell B.G.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL132779; CAB60015.1; -
DR PIR: T39116; T39116.
DR GeneDB_Sfombe; SPAC869.05C; -

```

```

DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008271; F:sulfate porter activity; IEA.
DR GO; GO:0008272; P:sulfate transport; IEA.
DR InterPro; IPR002645; STAS.
DR InterPro; IPR001902; Sulph_transp.
DR Pfam; PF01740; STAS; 1.
DR Pfam; PF00916; Sulfate_transp; 1.
DR TIGRFAMs; TIGR00815; sulP; 1.
DR PROSITE; PSS0801; STAS; 1.
SQ SEQUENCE 840 AA; 93517 MW; ED4833E162B69077 CRC64;

Query Match
Best Local Similarity 53.8%; Score 42; DB 3; Length 840;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 KRDNHYSKNP 12
DB 37 QNDHNDYTON 46

RESULT 15
Q9N4X7 PRELIMINARY; PRT; 378 AA.
AC Q9N4X7;
DT 01-OCT-2000 (TRENBLREL. 15, Created)
DT 01-OCT-2000 (TRENBLREL. 15, Last sequence update)
DT 01-OCT-2003 (TRENBLREL. 25, Last annotation update)
DE Hypothetical protein.
GN Y45G12C.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Bradshaw H., Graves T., Fronick B.;
RT "The sequence of C. elegans cosmid Y45G12C.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC006769; AAF60579.1; -
DR WormPep; Y45G12C.1; CE21936.
KW Hypothetical protein.
SQ SEQUENCE 378 AA; 44406 MW; E0B8EAD2F8507204 CRC64;

Query Match
Best Local Similarity 52.6%; Score 41; DB 5; Length 378;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 RDHNDYSKNP 13
DB 362 RDHNDYFMP 371

Search completed: March 22, 2004, 06:59:08
Job time : 5.81712 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:59:34 ; Search time 2.86381 Seconds

(without alignments)  
1265.926 Million cell updates/sec

Title: US-09-662-293-1

Perfect score: 78

Sequence: 1 SIKRDHNDYXKPM 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1049977 seqs, 258955339 residues

Total number of hits satisfying chosen parameters: 1049977

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA.\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78	100.0	14	US-10-218-743-1	Sequence 1, Appl
2	78	100.0	33	US-10-218-743-24	Sequence 24, Appl
3	78	100.0	536	US-10-218-743-21	Sequence 21, Appl
4	78	100.0	555	US-10-218-743-15	Sequence 15, Appl
5	78	100.0	555	US-10-218-743-18	Sequence 18, Appl
6	83.3	83.3	490	US-10-218-743-41	Sequence 41, Appl
7	83.3	83.3	509	US-10-218-743-35	Sequence 35, Appl
8	83.3	83.3	509	US-10-218-743-38	Sequence 38, Appl
9	83.3	83.3	355	US-10-424-599-259092	Sequence 259092,
10	42	53.8	587	US-10-259-165-134	Sequence 134, App
11	42	53.8	587	US-10-259-165-464	Sequence 464, App
12	40	51.3	61	US-10-424-599-262038	Sequence 262038,
13	40	51.3	383	US-10-282-122A-47229	Sequence 47229, A
14	39.5	50.6	442	US-09-881-752A-148	Sequence 148, App
15	39	50.0	317	US-10-282-122A-48754	Sequence 48754, A

16	39	50.0	507	12	US-10-425-114-72824	Sequence 72824, A
17	39	50.0	1418	15	US-10-369-493-22315	Sequence 22315, A
18	38.5	49.4	567	15	US-10-369-493-22125	Sequence 22125, A
19	38	48.7	154	12	US-10-424-599-187272	Sequence 187272,
20	38	48.7	211	12	US-10-282-122A-71431	Sequence 71431, A
21	38	48.7	557	12	US-10-282-122A-78546	Sequence 78546, A
22	38	48.7	645	12	US-10-282-122A-71875	Sequence 71875, A
23	38	48.7	764	9	US-09-866-582-41	Sequence 41, Appl
24	38	48.7	784	9	US-09-867-805-7	Sequence 7, Appl
25	37	47.4	17	10	US-09-592-665-48	Sequence 48, Appl
26	37	47.4	50	9	US-09-864-761-42504	Sequence 42504, A
27	37	47.4	171	12	US-10-424-599-251571	Sequence 251571,
28	37	47.4	220	15	US-10-369-493-11025	Sequence 11025, A
29	37	47.4	302	12	US-10-424-599-234827	Sequence 234827,
30	37	47.4	330	12	US-10-425-114-65174	Sequence 65174, A
31	37	47.4	359	13	US-10-029-756-2	Sequence 2, Appl
32	37	47.4	359	14	US-10-340-779A-12	Sequence 12, Appl
33	37	47.4	359	15	US-10-369-493-2766	Sequence 2766, Ap
34	37	47.4	386	15	US-10-412-699B-720	Sequence 720, App
35	37	47.4	386	15	US-10-374-780A-2526	Sequence 2526, Ap
36	37	47.4	507	9	US-09-910-087-19	Sequence 19, Appl
37	37	47.4	509	9	US-09-910-087-21	Sequence 21, Appl
38	37	47.4	533	15	US-10-369-493-12632	Sequence 12632, A
39	37	47.4	589	12	US-10-424-599-195059	Sequence 195059,
40	37	47.4	775	15	US-10-108-260A-2844	Sequence 2844, Ap
41	37	47.4	928	12	US-10-425-114-39886	Sequence 39886, A
42	37	47.4	1163	10	US-09-949-029-142	Sequence 142, App
43	37	47.4	1174	15	US-10-094-749-1911	Sequence 1911, Ap
44	37	47.4	2835	9	US-09-885-535-4	Sequence 4, Appl
45	36.5	46.8	248	12	US-10-282-122A-67090	Sequence 67090, A

## ALIGNMENTS

RESULT 1  
US-10-218-743-1  
; Sequence 1, Application US/10218743  
; Publication No. US20030096779A1  
GENERAL INFORMATION:  
; APPLICANT: McCall, Catherine A.  
; APPLICANT: Hunter, Shirley Wu  
; APPLICANT: Weber, Eric R.  
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: AL-2-C3  
; CURRENT APPLICATION NUMBER: US/10/218, 743  
; CURRENT FILING DATE: 2002-08-13  
; PRIOR APPLICATION NUMBER: US/09/292,225  
; PRIOR FILING DATE: 1998-04-15  
; PRIOR APPLICATION NUMBER: 60/098, 909  
; PRIOR FILING DATE: 1998-09-02  
; PRIOR APPLICATION NUMBER: 60/085, 295  
; PRIOR FILING DATE: 1998-05-13  
; PRIOR APPLICATION NUMBER: 60/098, 565  
; PRIOR FILING DATE: 1998-04-17  
; PRIOR APPLICATION NUMBER: 09/062, 013  
; PRIOR FILING DATE: 1998-04-17  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 14  
; TYPE: PRT  
; ORGANISM: Dermatophagoides farinae  
US-10-218-743-1  
Query Match 100.0%; Score 78; DB 14; Length 14;  
Best Local Similarity 100.0%; Pred. No. 1.7e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 1 SIKRDHNDYXKPM 14  
DB 1 SIKRDHNDYXKPM 14

```
RESULT 2
US-10-218-743-24
; Sequence 24, Application US/10218743
; Publication No. US20030096779A1
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/10/218,743
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US/09/292,225
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: 60/098,909
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/085,295
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/098,565
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: 09/062,013
; PRIOR FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 33
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
; FEATURE:
; OTHER INFORMATION: At locations, 18, 28, 31 and 32, Xaa = any amino
; OTHER INFORMATION: acid
US-10-218-743-24

Query Match          100.0%; Score 78; DB 14; Length 33;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SIKRDHNDYSKNPM 14
        |||||
        1 SIKRDHNDYSKNPM 14

RESULT 3
US-10-218-743-21
; Sequence 21, Application US/10218743
; Publication No. US20030096779A1
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/10/218,743
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US/09/292,225
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: 60/098,909
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/085,295
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/098,565
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: 09/062,013
; PRIOR FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 536
; TYPE: PRT
```

```
; ORGANISM: Dermatophagoides farinae
US-10-218-743-21

Query Match          100.0%; Score 78; DB 14; Length 536;
Best Local Similarity 100.0%; Pred. No. 6.2e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SIKRDHNDYSKNPM 14
        |||||
        1 SIKRDHNDYSKNPM 14

Db      20 SIKRDHNDYSKNPM 33

RESULT 4
US-10-218-743-15
; Sequence 15, Application US/10218743
; Publication No. US20030096779A1
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/10/218,743
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US/09/292,225
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: 60/098,909
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/085,295
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/098,565
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: 09/062,013
; PRIOR FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-10-218-743-15

Query Match          100.0%; Score 78; DB 14; Length 555;
Best Local Similarity 100.0%; Pred. No. 6.5e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 SIKRDHNDYSKNPM 14
        |||||
        20 SIKRDHNDYSKNPM 33

RESULT 5
US-10-218-743-18
; Sequence 18, Application US/10218743
; Publication No. US20030096779A1
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/10/218,743
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US/09/292,225
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: 60/098,909
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/085,295
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 09/062,013
; PRIOR FILING DATE: 1998-04-17
```

PRIOR APPLICATION NUMBER: 09/062,013  
PRIOR FILING DATE: 1998-04-17  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 18  
LENGTH: 555  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-10-218-743-18

Query Match 100.0%; Score 78; DB 14; Length 555;  
Best Local Similarity 100.0%; Pred. No. 6.5e-05;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KRDNHNSKNPM 14  
DB 20 SKRDNHNSKNPM 33

RESULT 6  
US-10-218-743-41  
Sequence 41, Application US/10218743  
Publication No. US20030096779A1  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley Wu  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/10/218,743  
CURRENT FILING DATE: 2002-08-13  
PRIOR APPLICATION NUMBER: US/09/292,225  
PRIOR FILING DATE: 1999-04-15  
PRIOR APPLICATION NUMBER: 60/098,909  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/085,295  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/098,565  
PRIOR FILING DATE: 1998-04-17  
PRIOR APPLICATION NUMBER: 09/062,013  
PRIOR FILING DATE: 1998-04-17  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 41  
LENGTH: 490  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-10-218-743-41

Query Match 83.3%; Score 65; DB 14; Length 490;  
Best Local Similarity 91.7%; Pred. No. 0.0094;  
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 KRDNHNSKNPM 14  
DB 3 KRDNHNSKNPM 14

RESULT 7  
US-10-218-743-35  
Sequence 35, Application US/10218743  
Publication No. US20030096779A1  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley Wu  
APPLICANT: Weber, Eric R.  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/10/218,743  
CURRENT FILING DATE: 2002-08-13  
PRIOR APPLICATION NUMBER: US/09/292,225

PRIOR FILING DATE: 1999-04-15  
PRIOR APPLICATION NUMBER: 60/098,909  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/085,295  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/098,565  
PRIOR FILING DATE: 1998-04-17  
PRIOR APPLICATION NUMBER: 09/062,013  
PRIOR FILING DATE: 1998-04-17  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 35  
LENGTH: 509  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-10-218-743-35

Query Match 83.3%; Score 65; DB 14; Length 509;  
Best Local Similarity 91.7%; Pred. No. 0.0098;  
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 KRDNHNSKNPM 14  
DB 22 KRDNHNSKNPM 33

RESULT 8  
US-10-218-743-38  
Sequence 38, Application US/10218743  
Publication No. US20030096779A1  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley Wu  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/10/218,743  
CURRENT FILING DATE: 2002-08-13  
PRIOR APPLICATION NUMBER: US/09/292,225  
PRIOR FILING DATE: 1999-04-15  
PRIOR APPLICATION NUMBER: 60/098,909  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/085,295  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/098,565  
PRIOR FILING DATE: 1998-04-17  
PRIOR APPLICATION NUMBER: 09/062,013  
PRIOR FILING DATE: 1998-04-17  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 38  
LENGTH: 509  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-10-218-743-38

Query Match 83.3%; Score 65; DB 14; Length 509;  
Best Local Similarity 91.7%; Pred. No. 0.0098;  
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 KRDNHNSKNPM 14  
DB 22 KRDNHNSKNPM 33

RESULT 9  
US-10-424-599-259092  
Sequence 259092, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K

RESULT 11  
US-10-259-165-464  
; Sequence 464, Application US/10259165  
; Publication No. US20030135888A1

## RESULT 13

```
US-10-282-122A-47229
; Sequence 47229, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,948
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47229
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-10-282-122A-47229

Query Match          51.3%; Score 40; DB 12; Length 383;
Best Local Similarity 70.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      2 IKRDNDYSK 11
      |||:|:|:|
      172 IKRDNSDFSK 181

Db      172 IKRDNSDFSK 181

RESULT 14
US-09-881-752A-148
; Sequence 148, Application US/09881752A
; Patent No. US20020115078A1
; GENERAL INFORMATION:
; APPLICANT: Kleanchous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Comen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020115078A1 Helicobacter Polypeptides in the
; FILE REFERENCE: 06133/041002
; CURRENT APPLICATION NUMBER: US/09/881,752A
```

```
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/833,457
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 148
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-881-752A-148

Query Match          50.6%; Score 39.5; DB 9; Length 442;
Best Local Similarity 44.4%; Pred. No. 1.9e+02;
Matches 8; Conservative 4; Mismatches 1; Indels 5; Gaps 1;

QY      2 IKRDH----DYKQPM 14
      |||:|:|:|:|
      310 IKRDHNSKIDYKPMPI 327

Db      310 IKRDHNSKIDYKPMPI 327

RESULT 15
US-10-282-122A-48754
; Sequence 48754, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48754
; LENGTH: 317
; TYPE: PRT
; ORGANISM: Bacteroides fragilis
US-10-282-122A-48754

Query Match          50.0%; Score 39; DB 12; Length 317;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

Mon Mar 22 11:57:24 2004

us-09-662-293-1.rapb

Page 6

QY 1 SIKKDNDY 9  
:|:|:|  
Db 209 NIKKDHDY 217

Search completed: March 22, 2004, 07:45:48  
Job time : 3.86381 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:30:23 ; Search time 4.25661 Seconds

(without alignments)  
929.256 Million cell updates/sec

Title: US-09-662-293-1

Perfect score: 78

Sequence: 1 SIRKDNNDYSKPM 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 262547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_293Jan04.\*

1: geneseqp1980s.\*  
2: geneseqp1990s.\*  
3: geneseqp2000s.\*  
4: geneseqp2001s.\*  
5: geneseqp2002s.\*  
6: geneseqp2003as.\*  
7: geneseqp2003bs.\*  
8: geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78	100.0	14	AAV52510	House dus
2	78	100.0	14	AAU96314	Der HMW-m
3	78	100.0	33	AAV52527	House dus
4	78	100.0	33	AAU96331	Der HMW-m
5	78	100.0	536	AAV52525	House dus
6	78	100.0	536	AAU96329	Der HMW-m
7	78	100.0	555	AAV52523	House dus
8	78	100.0	555	AAU96327	Der HMW-m
9	78	100.0	555	AAU96328	Der HMW-m
10	65	83.3	490	AAV52535	D. pteron
11	65	83.3	490	AAU96339	Der HMW-m
12	65	83.3	509	AAV52533	D. pteron
13	65	83.3	509	AAU96337	Der HMW-m
14	65	83.3	509	AAU96338	Der HMW-m
15	45	57.7	445	AAU96376	Human pro
16	40	51.3	143	AAU46683	Propionib
17	40	51.3	143	ABM43402	Propionib
18	40	51.3	348	AAV19899	B. burgdo
19	40	51.3	383	AAV19898	B. burgdo
20	40	51.3	383	ABU19305	Protein e
21	40	51.3	445	ADC31765	Human nov
22	39.5	50.6	442	AAV98274	H. pylori
23	39.5	50.6	442	ABU51548	Helicobac
24	39	50.0	317	ABU20830	Protein e
25	39	50.0	353	AAU91896	Sequence

26	39	50.0	367	6	ABU01816
27	39	50.0	374	4	ABG04285
28	39	50.0	444	1	AAU80669
29	39	50.0	444	2	AAV05371
30	39	50.0	444	5	ABR97582
31	39	50.0	655	6	ABR53677
32	39	50.0	1418	6	ABR53853
33	38	48.7	80	4	AAU86958
34	38	48.7	150	5	ABP40254
35	38	48.7	183	3	ABU16612
36	38	48.7	211	6	ABU43507
37	38	48.7	382	4	ABR63478
38	38	48.7	444	6	ABM70283
39	38	48.7	533	4	ABR67019
40	38	48.7	533	4	ABR67020
41	38	48.7	533	4	ABR60202
42	38	48.7	557	6	ABU50622
43	38	48.7	645	6	ABU43951
44	38	48.7	764	5	AAU81523
45	38	48.7	784	3	AAV82948

## ALIGNMENTS

RESULT 1  
ID AAV52510 standard; peptide; 14 AA.  
XX  
XX AAV52510;  
XX  
XX 22-FEB-2000 (first entry)  
XX  
XX House dust mite allergen protein (map) A/B N-terminal fragment.  
DE  
XX  
XX Mite allergen protein; map; high molecular weight; HMW-map; allergy;  
KW house dust mite; Igs; immunoglobulin E; allergen; map; map;  
KW hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
KW canine; veterinary; antibody; vaccine; immunisation.  
XX  
XX Dermatophagoides farinae.  
XX  
XX WO954349-A2.  
XX  
XX 28-OCT-1999.  
XX  
XX 16-APR-1999; 99WC-US008524.  
XX  
XX 17-APR-1998; 98US-00062013.  
XX  
XX 13-MAY-1998; 98US-0085295P.  
XX  
XX 02-SEP-1998; 98US-0098909P.  
XX  
XX (HESK-) HESKA CORP.  
XX  
XX Mccall CA, Hunter SW, Weber ER;  
XX WPI; 2000-052700/04.  
XX  
XX Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
XX used to modify an animal's hypersensitivity to mite allergens.  
XX  
XX Claim 3; Page 69; 154pp; English.  
XX  
XX Sequences AAV52510-Y52522 represent proteolytic fragments of  
XX Dermatophagoides farinae high molecular weight mite allergen protein (HMW  
XX -map) composition. The HMW-map composition was isolated from a D. farinae  
XX homogenate by gel filtration, with each fraction being analysed for the  
XX presence of proteins that bound to Igs present in mite-allergic dog  
XX antisera. The HMW-map composition comprises mapA (a 109 kD protein) and  
XX mapB (98 kD). Mite allergenic proteins and peptides, and nucleic acids  
XX encoding them, may be used in therapeutic compositions to modify an  
XX animal's hypersensitivity reaction to mite allergens. Animals that may be  
XX treated include mammals and birds, especially felines, canines, equines,

CC humans, other pets, and work or domestic animals. The proteins or  
 CC fragments may also be used to diagnose allergies via a skin test. The  
 CC proteins and peptides can also be used to raise antibodies, which have a  
 CC variety of potential uses. For example, they can be used as vaccines to  
 CC passively immunise animals against dust mite hypersensitivity, as  
 CC positive controls in test kits and as tools to recover desired dust mite  
 CC allergens from a mixture of proteins

XX Sequence 14 AA;

SO Query Match 100.0%; Score 78; DB 3; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 8,7e-07;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIKRDNDYKNPM 14  
 |||||  
 1 SIKRDNDYKNPM 14

RESULT 2  
 ID AAU96314 standard; peptide; 14 AA.

XX AAU96314;  
 XX 15-JUL-2002 (first entry)  
 DT Der HMW-map polypeptide #1.

XX Der HMW-map; American house dust mite; anti-allergic; mite; IGE;  
 KW mite allergenic protein; immunoglobulin E; hypersensitivity;  
 KM immunocomplex formation.

XX Dermatophagoides farinae.

XX WO200222807-A2.

XX 21-MAR-2002.

XX 14-SEP-2001; 2001MO-US028730.

XX 14-SEP-2000; 2000US-00662293.

XX (HESK-) HESKA CORP.

XX McCall CA, Hunter SW, Weber ER;

XX WPI; 2002-351888/38.

PT New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HMW-map protein, useful as a vaccine for treating mite allergy.

XX Claim 12; Page 70; 161pp; English.

CC The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic  
 CC acid. The Der HMW-map protein is useful for eliciting an immune response  
 CC against Der HMW-map protein. The protein or a reagent comprising a non-  
 CC prolineaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitizing a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig) E or Der HMW-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HMW-map polypeptides of the invention

XX Sequence 14 AA;

SO Query Match 100.0%; Score 78; DB 5; Length 14;

Best Local Similarity 100.0%; Pred. No. 8,7e-07;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SIKRDNDYKNPM 14  
 |||||  
 1 SIKRDNDYKNPM 14

RESULT 3  
 ID AAU96327 standard; peptide; 33 AA.

XX AAU96327;

DT 22-FEB-2000 (first entry)

XX House dust mite allergen protein (map) fragment, SEQ ID 24.

XX Mite allergen protein; map; high molecular weight; HMW-map; allergy;

KW house dust mite; IGE; immunoglobulin E; allergen;  
 KW hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
 KW canine; veterinary; antibody; vaccine; immunisation.

XX Dermatophagoides farinae.

XX Key Location/Qualifiers

FT Misc-difference 18 /label= Xaa

FT /note= "Xaa = any amino acid"

FT Misc-difference 28 /label= Xaa

FT /note= "Xaa = any amino acid"

FT Misc-difference 31 /label= Xaa

FT /note= "Xaa = any amino acid"

XX WO9954349-A2.

XX 28-OCT-1999.

XX 16-APR-1999; 99MO-US008524.

XX 17-APR-1998; 98US-00062013.

XX 13-MAY-1998; 98US-0085295P.

XX 02-SEP-1998; 98US-0098909P.

XX (HESK-) HESKA CORP.

XX McCall CA, Hunter SW, Weber ER;

XX WPI; 2000-052700/04.

PT Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
 PT used to modify an animals' hypersensitivity to mite allergens.

XX Claim 3; Page 78; 154pp; English.

CC This sequence represents an N-terminal proteolytic fragment of the  
 CC Dermatophagoides farinae high molecular weight mite allergen protein (HMW  
 CC -map) composition. The HMW-map composition was isolated from a D. farinae  
 CC homogenate by gel filtration, with each fraction being analysed for the  
 CC presence of proteins that bound to IgE present in mite-allergic dog  
 CC antisera. Mite allergenic proteins and peptides, and nucleic acids  
 CC encoding them, may be used in therapeutic compositions to modify an  
 CC animal's hypersensitivity reaction to mite allergens. Animals that may be  
 CC treated include mammals and birds, especially felines, canines, equines,  
 CC humans, other pets, and work or domestic animals. The proteins or  
 CC fragments may also be used to diagnose allergies via a skin test. The  
 CC proteins and peptides can also be used to raise antibodies, which have a  
 CC variety of potential uses. For example, they can be used as vaccines to



CC passively immunise animals against dust mite hypersensitivity, as  
 CC positive controls in test kits and as tools to recover desired dust mite  
 CC allergens from a mixture of proteins

XX Sequence 33 AA;

Query Match 100.0%; Score 78; DB 3; Length 33;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIKRDNDYSKNPM 14  
 |||||  
 1 SIKRDNDYSKNPM 14

RESULT 4  
 AAU96331  
 ID AAU96331 standard; peptide; 33 AA.

XX AAU96331;

DT 15-JUL-2002 (first entry)

DE Der HMW-map polypeptide #18.

XX Der HMW-map; American house dust mite; anti-allergic; mite; IGE;  
 KW mite allergenic protein; immunoglobulin E; hypersensitivity;  
 KM immunocomplex formation.

OS Dermatophagoides farinae.

XX WO200222807-A2.

PN 21-MAR-2002.

XX 14-SEP-2001; 2001WO-US028730.

PF 14-SEP-2000; 2000US-00662293.

PR (HESK-) HESKA CORP.

PA Mccall CA, Hunter SW, Weber ER;

PI WPI; 2002-351888/38.

PT New mite allergenic protein isolated from Dermatophagoides, designated  
 Der HMW-map protein, useful as a vaccine for treating mite allergy.

PS Claim 12; Page 79; 161pp; English.

XX The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic  
 CC acid. The Der HMW-map protein is useful for eliciting an immune response  
 CC against Der HMW-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitising a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (IgE) or Der HMW-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HMW-map polypeptides of the invention

XX Sequence 33 AA;

Query Match 100.0%; Score 78; DB 5; Length 33;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIKRDNDYSKNPM 14

Db |||||  
 1 SIKRDNDYSKNPM 14

RESULT 5  
 AAU52525  
 ID AAU52525 standard; protein; 536 AA.

XX AAU52525;

DT 22-FEB-2000 (first entry)

DE House dust mite (D. farinae) mite allergen protein (map) Pderf98-536.

XX Mite allergen protein; map; high molecular weight; HMW-map; allergy;  
 KW house dust mite; IGE; immunoglobulin E; allergen; map; mapB;  
 KM hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
 KW canine; veterinary; antibody; vaccine; immunisation.

XX Dermatophagoides farinae.

XX WO9954349-A2.

PD 28-OCT-1999.

PF 16-APR-1999; 99WO-US008524.

XX 17-APR-1998; 98US-00062013.

PR 13-MAY-1998; 98US-0085295P.

PR 02-SEP-1998; 98US-0098909P.

PA (HESK-) HESKA CORP.

PI Mccall CA, Hunter SW, Weber ER;

DR WPI; 2000-052700/04.

DR N-PSDB; AA238579, AA238580.

PT Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
 used to modify an animal's hypersensitivity to mite allergens.

PS Claim 3; Page 125-127; 154pp; English.

XX This sequence represents Dermatophagoides farinae mite allergen protein  
 CC (map) Pderf98-536, the mature form of Pderf98-555 (AAU52525). Pderf98-536  
 CC has a molecular weight of 98 kD, comprising 536 amino acids, and is a  
 CC component of the Dermatophagoides farinae high molecular weight mite  
 CC allergen protein (HMW-map) composition. The HMW-map composition was  
 CC isolated from a D. farinae homogenate by gel filtration, with each  
 CC fraction being analysed for the presence of proteins that bound to IgE  
 CC present in mite-allergic dog antisera. Mite allergenic proteins and  
 CC peptides, and nucleic acids encoding them, may be used in therapeutic  
 CC compositions to modify an animal's hypersensitivity reaction to mite  
 CC allergens. Animals that may be treated include mammals and birds,  
 CC especially felines, canines, equines, humans, other pets, and work or  
 CC domestic animals. The proteins or fragments may also be used to diagnose  
 CC allergies via a skin test. The proteins and peptides can also be used to  
 CC raise antibodies, which have a variety of potential uses. For example,  
 CC they can be used as vaccines to passively immunise animals against dust  
 CC mite hypersensitivity, as positive controls in test kits and as tools to  
 CC recover desired dust mite allergens from a mixture of proteins

XX Sequence 536 AA;

Query Match 100.0%; Score 78; DB 3; Length 536;  
 Best Local Similarity 100.0%; Pred. No. 6.4e-05;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIKRDNDYSKNPM 14  
 |||||  
 1 SIKRDNDYSKNPM 14

```

RESULT 6
AAU96329 AAU96329 standard; protein; 536 AA.
ID AAU96329
AC AAU96329;
DT 15-JUL-2002 (first entry)
DE Der HMW-map polypeptide #16.
DS
XX Der HMW-map; American house dust mite; antiallergic; mite; IgE;
KW mite allergenic protein; immunoglobulin E; hypersensitivity;
KV immunocomplex formation.
XX
OS Dermatophagoides farinae.
PN WO20022807-A2.
PP 21-MAR-2002.
PR 14-SEP-2001; 2001WO-US028730.
PX 14-SEP-2000; 2000US-00662293.
PY (HESK-) HESKA CORP.
PI McCall CA, Hunter SW, Weber ER;
PT WPI; 2002-351868/38.
PS N-PsDB; ABX69575.
XX New mite allergenic protein isolated from Dermatophagoides, designated
XX Der HMW-map protein, useful as a vaccine for treating mite allergy.
XX Claim 12; Page 125-127; 16pp; English.
XX
XX The invention relates to an isolated mite allergenic protein of
CC dermatophagoides, designated Der HMW-map protein, and its related nucleic
CC acid. The Der HMW-map protein is useful for eliciting an immune response
CC against Der HMW-map protein. The protein or a reagent comprising a non-
CC proteinaceous epitope is useful for identifying an animal (e.g., dog,
CC cat) susceptible to or having an allergic response to a mite. A
CC therapeutic composition is useful for desensitizing a host animal to an
CC allergic response to a mite. The DNA and protein can be used in the
CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition
CC of immunoglobulin (Ig) E or Der HMW-map protein activity associated with a
CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting
CC binding of proteins to IgE, to prevent immunocomplex formation, thus
CC reducing hypersensitivity responses to mite allergens, and as vaccines
CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342
CC represent Der HMW-map polypeptides of the invention
CC
SQ Sequence 536 AA;
QY Query Match 100.0%; Score 78; DB 5; Length 536;
Best Local Similarity 100.0%; Pred. No. 6;4e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0
DB 1 SIXRDNDYSKNPM 14
1 SIXRDNDYSKNPM 14
RESULT 7
AAU96329 ID AAU96329 standard; protein; 555 AA.
AC AAU96329;
DT 22-FEB-2000 (first entry)
XX House dust mite (D. farinae) mite allergen protein (map) PDerf98-555.
XX

```

KM	Mite allergen protein; map; high molecular weight; HMW-map; allergy;
KM	house dust mite; IGE; immunoglobulin E; allergen; mapA; mapB;
KM	hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;
KM	canine; veterinary; antibody; vaccine; immunisation.
XX	
OS	Dermatophagoides farinae.
XX	
FH	Key Location/Qualifiers
FT	Peptide 1..19 /note= "Signal peptide"
FT	Protein 20..555
FT	/note= "Mature Pderf98-555"
XX	
BN	WO954349-A2.
PD	28-OCT-1999.
XX	
PF	16-APR-1999; 99WO-US008524.
PR	17-APR-1998; 98US-00062013.
PR	13-MAY-1998; 98US-0085285P.
XX	02-SEP-1998; 98US-0098909P.
PA	(HESK-) HESKA CORP.
XX	
PI	Mccall CA, Hunter SW, Weber ER;
DR	WP1; 2000-052700/04.
XX	N-PSDB; AAZ38575, AAZ38576, AAZ38577, AAZ38578.
PT	Novel high molecular weight Dermatophagoides nucleic acid polypeptides used to modify an animals' hypersensitivity to mite allergens.
XX	
PS	Claim 3; Page 111-113; 154pp; English.
XX	
CC	This sequence represents Dermatophagoides farinae mite allergen protein (map). Pderf98-555. Pderf98-555 has a molecular weight of 98 kD, comprising 555 amino acids, and is a component of the Dermatophagoides farinae high molecular weight mite allergen protein (HMW-map) composition. The HMW-map composition was isolated from a D. Farinae homogenate by gel filtration, with each fraction being analysed for the presence of proteins that bound to IGE present in mite-allergic dog antisera. Mite allergenic proteins and peptides, and nucleic acids encoding them, may be used in therapeutic compositions to modify an animal's hypersensitivity reaction to mite allergens. Animals that may be treated include mammals and birds, especially felines, canines, equines, humans, other pets, and work or domestic animals. The proteins or fragments may also be used to diagnose allergies via a skin test. The proteins and peptides can also be used to raise antibodies, which have a variety of potential uses. For example, they can be used as vaccines to passively immunise animals against dust mite hypersensitivity, as positive controls in test kits and as tools to recover desired dust mite allergens from a mixture of proteins
XX	
SQ	Sequence 555 AA:
QY	Query Match 100.0%; Score 78; DB 3; Length 555; Best Local Similarity 100.0%; Pred. No. 6,7e-05; Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0.  1 SIKRDNDYSKNPM 14       14 SIKRDNDYSKNPM 33
ID	AAU96327 standard; protein; 555 AA.
AC	AAU96327;
XX	
XX	
DT	15-JUL-2002 (first entry)
XX	

DE Der HMW-map polypeptide #14.  
 XX  
 KW Der HMW-map; American house dust mite; antiallergic; mite; IgE;  
 KW mite allergenic protein; immunoglobulin E; hypersensitivity;  
 XX immunocomplex formation.  
 OS Dermatophagoides farinae.  
 XX  
 PN WO200222807-A2.  
 XX  
 PD 21-MAR-2002.  
 XX  
 PP 14-SEP-2001; 2001WO-US028730.  
 XX  
 PR 14-SEP-2000; 2000US-00662293.  
 XX  
 PA (HESK-) HESKA CORP.  
 XX  
 PI McCall CA, Hunter SM, Weber ER;  
 XX  
 DR WPI; 2002-351888/38.  
 DR N-PSDB; ABK69571.  
 XX  
 PT New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HMW-map protein, useful as a vaccine for treating mite allergy.  
 XX  
 PS Claim 12; Page 114-116; 161pp; English.  
 XX  
 CC The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic  
 CC acid. The Der HMW-map protein is useful for eliciting an immune response  
 CC against Der HMW-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitizing a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig)E or Der HMW-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HMW-map polypeptides of the invention  
 XX  
 SO Sequence 555 AA;  
 XX  
 Query Match 100.0%; Score 78; DB 5; Length 555;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-05;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SIXRDNDYSKNPM 14  
 |||||  
 DB 20 SIXRDNDYSKNPM 33  
 |||||  
 RESULT 9  
 AAU96328  
 ID AAU96328 standard; protein; 555 AA.  
 XX  
 AC AAU96328;  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX  
 DE Der HMW-map polypeptide #15.  
 XX  
 KW Der HMW-map; American house dust mite; antiallergic; mite; IgE;  
 KW mite allergenic protein; immunoglobulin E; hypersensitivity;  
 XX immunocomplex formation.  
 XX  
 OS Dermatophagoides farinae.  
 XX  
 PN WO200222807-A2.  
 XX

PD 21-MAR-2002.  
 XX  
 PP 14-SEP-2001; 2001WO-US028730.  
 XX  
 PR 14-SEP-2000; 2000US-00662293.  
 XX  
 PA (HESK-) HESKA CORP.  
 XX  
 PI McCall CA, Hunter SM, Weber ER;  
 XX  
 DR WPI; 2002-351888/38.  
 DR N-PSDB; ABK69573.  
 XX  
 PT New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HMW-map protein, useful as a vaccine for treating mite allergy.  
 XX  
 PS Claim 12; Page 120-122; 161pp; English.  
 XX  
 CC The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic  
 CC acid. The Der HMW-map protein is useful for eliciting an immune response  
 CC against Der HMW-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitizing a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig)E or Der HMW-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HMW-map polypeptides of the invention  
 XX  
 SO Sequence 555 AA;  
 XX  
 Query Match 100.0%; Score 78; DB 5; Length 555;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-05;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SIXRDNDYSKNPM 14  
 |||||  
 DB 20 SIXRDNDYSKNPM 33  
 |||||  
 RESULT 10  
 AAU52535  
 ID AAU52535 standard; protein; 490 AA.  
 XX  
 AC AAU52535;  
 XX  
 DT 06-AUG-2003 (revised)  
 DT 22-FEB-2000 (first entry)  
 XX  
 DE D. pteronyssinus 98 kD mite allergen protein (map) PDerp98-490.  
 XX  
 KW Mite allergen protein; map; high molecular weight; HMW-map; allergy;  
 KW house dust mite; IgE; immunoglobulin E; allergen; map;  
 KW hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
 KW canine; veterinary; antibody; vaccine; immunisation.  
 XX  
 OS Dermatophagoides pteronyssinus.  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 115..117  
 FT Modified-site 240..242  
 FT Modified-site 240..242  
 XX  
 PN WO954349-A2.  
 XX  
 PD 28-OCT-1999.  
 XX

PF 16-APR-1999; 99WO-US008524.  
 XX 17-APR-1998; 98US-00062013.  
 PR 13-MAY-1998; 98US-0085295P.  
 PR 02-SEP-1998; 98US-0098909P.  
 XX (HESK-) HESKA CORP.  
 PA Mccall CA, Hunter SM, Weber ER;  
 PI MPI; 2000-052700/04.  
 DR N-PSDB; AAZ38589, AAZ38590.  
 XX Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
 PT used to modify an animals' hypersensitivity to mite allergens.  
 PS Claim 3; Page 147-149; 154pp; English.  
 XX This sequence represents Dermatophagoides pteronyssinus mite allergen  
 CC protein (map) Pderp98-490, the mature form of Pderp98-509. Pderp98-490  
 CC has a molecular weight of 98 kD, comprising 490 amino acids, and has a  
 CC high degree of homology with the D. farinae mature 98 kD allergen, mapB  
 CC (AAV52825). Nucleic acid molecules encoding Pderp98-490 were isolated  
 CC from a D. pteronyssinus cDNA library by hybridisation with a probe  
 CC encoding the D. farinae high molecular weight map (HWM-map) composition.  
 CC Mite allergenic proteins and peptides, and nucleic acids encoding them,  
 CC may be used in therapeutic compositions to modify an animal's  
 CC hypersensitivity reaction to mite allergens. Animals that may be treated  
 CC include mammals and birds, especially felines, canines, equines, humans,  
 CC other pets, and work or domestic animals. The proteins or fragments may  
 CC also be used to diagnose allergies via a skin test. The proteins and  
 CC peptides can also be used to raise antibodies, which have a variety of  
 CC potential uses. For example, they can be used as vaccines to passively  
 CC immunise animals against dust mite hypersensitivity, as positive controls  
 CC in test kits and as tools to recover desired dust mite allergens from a  
 CC mixture of proteins. (Updated on 06-AUG-2003 to correct OS field.)  
 CC  
 SQ Sequence 490 AA;  
 Query Match 83.3%; Score 65; DB 3; Length 490;  
 Best Local Similarity 91.7%; Pred. No. 0.011;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 KRDNHYSKNPM 14  
 |||||:|||||  
 3 KRDNHYSKNPM 14  
 Db 3 KRDNHYSKNPM 14  
 RESULT 11  
 AAU96339  
 ID AAU96339 standard; protein; 490 AA.  
 AC AAU96339;  
 XX 15-JUL-2002 (first entry)  
 DT  
 XX Der HWM-map polypeptide #26.  
 DE Der HWM-map polypeptide #26.  
 XX Der HWM-map; American house dust mite; anti-allergic; mite; IgE;  
 KW mite allergenic protein; immunoglobulin E; hypersensitivity;  
 KM immunocomplex formation.  
 XX  
 OS Dermatophagoides farinae.  
 XX  
 PN WO200222807-A2.  
 XX  
 PD 21-MAR-2002.  
 XX  
 PF 14-SEP-2001; 2001WO-US028730.  
 PR 13-MAY-1998; 98US-0085295P.  
 PR 14-SEP-2000; 2000US-00662293.  
 XX  
 PA (HESK-) HESKA CORP.

XX Mccall CA, Hunter SM, Weber ER;  
 PI MPI; 2002-351888/38.  
 DR N-PSDB; AAK69585.  
 XX New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HWM-map protein, useful as a vaccine for treating mite allergy.  
 XX  
 PS Claim 12; Page 144-146; 161pp; English.  
 XX The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HWM-map protein, and its related nucleic  
 CC acid. The Der HWM-map protein is useful for eliciting an immune response  
 CC against Der HWM-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitising a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HWM-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig)E or Der HWM-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HWM-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96334-AAU96342  
 CC represent Der HWM-map polypeptides of the invention  
 CC  
 SQ Sequence 490 AA;  
 Query Match 83.3%; Score 65; DB 5; Length 490;  
 Best Local Similarity 91.7%; Pred. No. 0.011;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 KRDNHYSKNPM 14  
 |||||:|||||  
 3 KRDNHYSKNPM 14  
 Db 3 KRDNHYSKNPM 14  
 RESULT 12  
 AAU52533  
 ID AAU52533 standard; protein; 509 AA.  
 AC AAU52533;  
 XX 06-AUG-2003 (revised)  
 DT 22-FEB-2000 (first entry)  
 XX  
 DE D. pteronyssinus 98 kD mite allergen protein (map) Pderp98-509.  
 XX  
 KW Mite allergen protein; map; high molecular weight; HWM-map; allergy;  
 KW house dust mite; IgE; immunoglobulin E; allergen; mapB;  
 KM hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
 KW canine; veterinary; antibody; vaccine; immunisation.  
 XX  
 OS Dermatophagoides pteronyssinus.  
 XX  
 PN WO954349-A2.  
 XX  
 PD 28-OCT-1999.  
 XX  
 PF 16-APR-1999; 99WO-US008524.  
 XX  
 PR 17-APR-1998; 98US-00062013.  
 PR 13-MAY-1998; 98US-0085295P.  
 PR 02-SEP-1998; 98US-0098909P.  
 XX  
 PA (HESK-) HESKA CORP.

XX Mccall CA, Hunter SW, Weber ER;  
 XX WPI; 2000-052700/04.  
 DR N-PSDB; AA238585, AA238586, AA238587, AA238588.  
 XX Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
 PT used to modify an animal's hypersensitivity to mite allergens.  
 XX Claim 3; Page 134-136; 154pp; English.  
 XX This sequence represents Dermatophagoides pteronyssinus mite allergen  
 CC protein (map) pDerp98-509. pDerp98-509 has a molecular weight of 98 kD,  
 CC comprising 509 amino acids, and has a high degree of homology with the D.  
 CC farinae 98 kD allergen, mapB (AAV52523). Nucleic acid molecules encoding  
 CC pDerp98-509 were isolated from a D. pteronyssinus cDNA library by  
 CC hybridisation with a probe encoding the D. farinae high molecular weight  
 CC map (Hmw-map) composition. Mite allergenic proteins and peptides, and  
 CC nucleic acids encoding them, may be used in therapeutic compositions to  
 CC modify an animal's hypersensitivity reaction to mite allergens. Animals  
 CC that may be treated include mammals and birds, especially felines, the  
 CC canines, equines, humans, other pets, and work or domestic animals. The  
 CC proteins or fragments may also be used to diagnose allergies via a skin  
 CC test. The proteins and peptides can also be used to raise antibodies,  
 CC which have a variety of potential uses. For example, they can be used as  
 CC vaccines to passively immunise animals against dust mite  
 CC hypersensitivity, as positive controls in test kits and as tools to  
 CC recover desired dust mite allergens from a mixture of proteins. (Updated  
 CC on 06-AUG-2003 to correct OS field.)  
 XX  
 SQ Sequence 509 AA;  
 Query Match 83.3%; Score 65; DB 3; Length 509;  
 Best Local Similarity 91.7%; Pred. No. 0.011; 0; Indels 0; Gaps 0;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 KRDNNDYSKNPM 14  
 |||||:|||||  
 Db 22 KRDNHNYSKNPM 33  
 RESULT 13  
 AAU96337  
 ID AAU96337 standard; protein; 509 AA.  
 XX  
 AC AAU96337;  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX  
 DE Der Hmw-map polypeptide #24.  
 XX  
 KW Der Hmw-map; American house dust mite; antiallergic; mite; IgE;  
 KM mite allergenic protein; immunoglobulin E; hypersensitivity;  
 XX immunocomplex formation.  
 XX  
 OS Dermatophagoides farinae.  
 XX  
 PN WO200222807-A2.  
 XX  
 PD 21-MAR-2002.  
 XX  
 PF 14-SEP-2001; 2001WO-US028730.  
 XX  
 PR 14-SEP-2000; 2000US-00662293.  
 XX  
 PA (HESK-) HESKA CORP.  
 XX  
 PI Mccall CA, Hunter SW, Weber ER;  
 XX  
 DR WPI; 2002-351888/38.  
 DR N-PSDB; ABK69581.  
 XX  
 PT New mite allergenic protein isolated from Dermatophagoides, designated

PT Der Hmw-map protein, useful as a vaccine for treating mite allergy.  
 XX  
 XX Claim 12; Page 134-136; 161pp; English.  
 PS  
 XX The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der Hmw-map protein, and its related nucleic  
 CC acid. The Der Hmw-map protein is useful for eliciting an immune response  
 CC against Der Hmw-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitising a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der Hmw-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig) E or Der Hmw-map protein activity associated with a  
 CC disease. Antibodies that bind to Der Hmw-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der Hmw-map polypeptides of the invention  
 XX  
 SQ Sequence 509 AA;  
 Query Match 83.3%; Score 65; DB 5; Length 509;  
 Best Local Similarity 91.7%; Pred. No. 0.011; 0; Indels 0; Gaps 0;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 KRDNNDYSKNPM 14  
 |||||:|||||  
 Db 22 KRDNHNYSKNPM 33  
 RESULT 14  
 AAU96338  
 ID AAU96338 standard; protein; 509 AA.  
 XX  
 AC AAU96338;  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX  
 DE Der Hmw-map polypeptide #25.  
 XX  
 KW Der Hmw-map; American house dust mite; antiallergic; mite; IgE;  
 KM mite allergenic protein; immunoglobulin E; hypersensitivity;  
 XX immunocomplex formation.  
 XX  
 OS Dermatophagoides farinae.  
 XX  
 PN WO200222807-A2.  
 XX  
 PD 21-MAR-2002.  
 XX  
 PF 14-SEP-2001; 2001WO-US028730.  
 XX  
 PR 14-SEP-2000; 2000US-00662293.  
 XX  
 PA (HESK-) HESKA CORP.  
 XX  
 PI Mccall CA, Hunter SW, Weber ER;  
 XX  
 DR WPI; 2002-351888/38.  
 DR N-PSDB; ABK69583.  
 XX  
 PT Der Hmw-map protein, useful as a vaccine for treating mite allergy.  
 XX  
 PS Claim 12; Page 139-141; 161pp; English.  
 XX  
 XX The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der Hmw-map protein, and its related nucleic  
 CC acid. The Der Hmw-map protein is useful for eliciting an immune response  
 CC against Der Hmw-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A

CC therapeutic composition is useful for desensitising a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (IgE or Der HMW-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HMW-map polypeptides of the invention  
 CC  
 SQ Sequence 509 AA;

Query March 83.3%; Score 65; DB 5; Length 509;  
 Best Local Similarity 91.7%; Pred. No. 0.011; 0; Indels 0; Gaps 0;  
 Matches 11; Conservative 1; Mismatches 0; Gaps 0;  
 QY 3 KRDNDYKSNPM 14  
 |||:|||||  
 Db 22 KRDNDYKSNPM 33

## RESULT 15

AA94676  
 ID AA94676 standard; protein; 445 AA.

AA94676;

26-JUN-2001 (first entry)

Human protein sequence SEQ ID NO:15624.

Human; primer; detection; diagnosis; antisense therapy; gene therapy.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUN-2000; 2000EP-00116126.

29-JUL-1999; 99JP-00248036.

27-AUG-1999; 99JP-00300253.

11-JAN-2000; 2000JP-00118776.

02-MAY-2000; 2000JP-00183767.

09-JUN-2000; 2000JP-00241899.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
 length cDNAs defined in the specification, and for the detection and/or  
 diagnosis of the abnormality of the proteins encoded by the full-length  
 cDNAs.

Claim 8; SEQ ID NO 15624; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesizing 5602 full-  
 length cDNAs defined in the specification. Where a primer set comprises:  
 (a) an oligo-dT primer and an oligonucleotide complementary to the  
 complementary strand of a polynucleotide which comprises one of the 5602  
 nucleotide sequences defined in the specification, where the  
 oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 of an oligonucleotide comprising a sequence complementary to the  
 complementary strand of a polynucleotide which comprises a 5'-end  
 sequence and an oligonucleotide comprising a sequence complementary to a  
 polynucleotide which comprises a 3'-end sequence, where the  
 oligonucleotide comprises at least 15 nucleotides and the combination of  
 the 5'-end sequence/3'-end sequence is selected from those defined in the

CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH1628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
 CC represent human amino acid sequences; and AAH1629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention  
 CC  
 SQ Sequence 445 AA;

Query March 57.7%; Score 45; DB 4; Length 445;  
 Best Local Similarity 58.3%; Pred. No. 31; 2; Indels 0; Gaps 0;  
 Matches 7; Conservative 3; Mismatches 0; Gaps 0;  
 QY 2 IRRDNDYKSNP 13  
 |||:|||||  
 Db 61 IRRDNDYKSNP 72

Search completed: March 22, 2004, 06:51:32  
 Job time : 7.25681 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:42:54 ; Search time 1.1284 Seconds  
(without alignments)  
640.518 Million cell updates/sec

Title: US-09-662-293-1  
Perfect score: 78  
Sequence: 1 SIXRDNDYXKPM 14

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgm2\_6/prodata/2/iaa/5A\_COMB.pep:\*  
2: /cgm2\_6/prodata/2/iaa/5B\_COMB.pep:\*  
3: /cgm2\_6/prodata/2/iaa/5A\_COMB.pep:\*  
4: /cgm2\_6/prodata/2/iaa/5B\_COMB.pep:\*  
5: /cgm2\_6/prodata/2/iaa/PCITUS\_COMB.pep:\*  
6: /cgm2\_6/prodata/2/iaa/backfill1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

No.	Score	Query Match	Length	DB	ID	Description
1	78	100.0	14	4	US-09-292-225-1	Sequence 1, Appli
2	78	100.0	33	4	US-09-292-225-34	Sequence 24, Appli
3	78	100.0	536	4	US-09-292-225-21	Sequence 21, Appli
4	78	100.0	555	4	US-09-292-225-15	Sequence 15, Appli
5	78	100.0	555	4	US-09-292-225-18	Sequence 18, Appli
6	65	83.3	490	4	US-09-292-225-41	Sequence 41, Appli
7	65	83.3	509	4	US-09-292-225-35	Sequence 35, Appli
8	65	83.3	509	4	US-09-292-225-38	Sequence 38, Appli
9	41	52.6	121	4	US-09-543-681A-7570	Sequence 7570, Appli
10	39	50.0	820	4	US-09-252-991A-25153	Sequence 25153, A
11	38	48.7	150	4	US-09-134-001C-5099	Sequence 5099, Ap
12	38	48.7	203	4	US-09-252-991A-24219	Sequence 24219, Ap
13	38	48.7	784	3	US-09-371-913A-7	Sequence 7, Appli
14	38	48.7	784	4	US-09-367-805-7	Sequence 7, Appli
15	38	48.7	947	4	US-09-719-190-2	Sequence 2, Appli
16	37	47.4	71	4	US-09-540-226-328	Sequence 2328, Ap
17	37	47.4	165	3	US-08-799-149C-5	Sequence 5, Appli
18	37	47.4	340	4	US-09-328-352-7764	Sequence 7764, Ap
19	37	47.4	352	1	US-08-552-142A-4	Sequence 4, Appli
20	37	47.4	352	1	US-08-910-973-4	Sequence 4, Appli
21	37	47.4	352	4	US-09-499-827-4	Sequence 4, Appli
22	37	47.4	352	5	PCT-US95-05741-4	Sequence 4, Appli
23	37	47.4	359	1	US-08-307-382-2	Sequence 2, Appli
24	37	47.4	359	1	US-08-366-779-2	Sequence 2, Appli
25	37	47.4	359	1	US-08-478-727-2	Sequence 2, Appli
26	37	47.4	359	1	US-08-473-508-2	Sequence 2, Appli
27	37	47.4	359	1	US-08-789-936-2	Sequence 2, Appli

## ALIGNMENTS

28	37	47.4	359	2	US-08-833-610-6	Sequence 6, Appl
29	37	47.4	359	3	US-08-834-033A-16	Sequence 16, Appl
30	37	47.4	359	4	US-08-934-254-2	Sequence 2, Appl
31	37	47.4	359	4	US-09-377-452-6	Sequence 6, Appl
32	37	47.4	359	4	US-09-685-775-2	Sequence 2, Appl
33	37	47.4	451	3	US-08-969-644-8	Sequence 8, Appl
34	37	47.4	451	3	US-08-444-189-8	Sequence 8, Appl
35	37	47.4	451	3	US-08-468-544-8	Sequence 8, Appl
36	37	47.4	507	3	US-08-860-635A-19	Sequence 19, Appl
37	37	47.4	507	4	US-09-281-476-19	Sequence 19, Appl
38	37	47.4	509	3	US-08-845-546-2	Sequence 2, Appl
39	37	47.4	509	3	US-08-860-635A-21	Sequence 21, Appl
40	37	47.4	509	4	US-09-281-476-21	Sequence 21, Appl
41	36	46.2	317	1	US-08-463-0908-9	Sequence 9, Appl
42	36	46.2	317	2	US-08-874-347-18	Sequence 18, Appl
43	36	46.2	317	2	US-09-093-522-18	Sequence 18, Appl
44	36	46.2	323	2	US-08-871-268A-20	Sequence 20, Appl
45	36	46.2	323	3	US-08-871-267B-28	Sequence 28, Appl

RESULT 1  
US-09-292-225-1  
Sequence 1, Application US/09292225  
Patent No. 6453686  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley Wu  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/09/292,225  
CURRENT FILING DATE: 1999-04-15  
EARLIER APPLICATION NUMBER: 60/098,909  
EARLIER FILING DATE: 1998-09-02  
EARLIER APPLICATION NUMBER: 60/085,295  
EARLIER FILING DATE: 1998-05-13  
EARLIER APPLICATION NUMBER: 60/098,565  
EARLIER FILING DATE: 1998-04-17  
EARLIER APPLICATION NUMBER: 09/062,013  
EARLIER FILING DATE: 1998-04-17  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 1  
LENGTH: 14  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-09-292-225-1

Query Match 100.0%; Score 78; DB 4; Length 14;  
Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIXRDNDYXKPM 14  
Db 1 SIXRDNDYXKPM 14

RESULT 2  
US-09-292-225-24  
Sequence 24, Application US/09292225  
Patent No. 6453686  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley Wu  
APPLICANT: Weber, Eric R.  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/09/292,225

CURRENT FILING DATE: 1999-04-15  
; EARLIER APPLICATION NUMBER: 60/098,909  
; EARLIER FILING DATE: 1998-09-02  
; EARLIER APPLICATION NUMBER: 60/085,295  
; EARLIER FILING DATE: 1998-05-13  
; EARLIER APPLICATION NUMBER: 60/098,565  
; EARLIER FILING DATE: 1998-04-17  
; EARLIER APPLICATION NUMBER: 09/062,013  
; EARLIER FILING DATE: 1998-04-17  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 24  
; LENGTH: 33  
; TYPE: PRT  
; ORGANISM: Dermatophagoides farinae  
; FEATURE:  
; OTHER INFORMATION: At locations, 18, 28, 31 and 32, Xaa = any amino  
; OTHER INFORMATION: acid  
US-09-292-225-24

Query Match 100.0%; Score 78; DB 4; Length 33;  
Best Local Similarity 100.0%; Pred. No. 4.3e-07;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIKRDNDYSKNPM 14  
DB 1 SIKRDNDYSKNPM 14

RESULT 3  
US-09-292-225-21

; Sequence 21, Application US/09292225  
; Patent No. 6455686  
; GENERAL INFORMATION:  
; APPLICANT: McCall, Catherine A.  
; APPLICANT: Hunter, Shirley Wu  
; APPLICANT: Weber, Eric R.  
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
; FILE REFERENCE: AL-2-C3  
; CURRENT APPLICATION NUMBER: US/09/292,225  
; CURRENT FILING DATE: 1999-04-15  
; EARLIER APPLICATION NUMBER: 60/098,909  
; EARLIER FILING DATE: 1998-09-02  
; EARLIER APPLICATION NUMBER: 60/085,295  
; EARLIER FILING DATE: 1998-05-13  
; EARLIER APPLICATION NUMBER: 60/098,565  
; EARLIER FILING DATE: 1998-04-17  
; EARLIER APPLICATION NUMBER: 09/062,013  
; EARLIER FILING DATE: 1998-04-17  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 21  
; LENGTH: 536  
; TYPE: PRT  
; ORGANISM: Dermatophagoides farinae  
US-09-292-225-21

Query Match 100.0%; Score 78; DB 4; Length 536;  
Best Local Similarity 100.0%; Pred. No. 9.2e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIKRDNDYSKNPM 14  
DB 1 SIKRDNDYSKNPM 14

RESULT 4  
US-09-292-225-15  
; Sequence 15, Application US/09292225  
; Patent No. 6455686  
; GENERAL INFORMATION:  
; APPLICANT: McCall, Catherine A.

; APPLICANT: Hunter, Shirley Wu  
; APPLICANT: Weber, Eric R.  
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
; FILE REFERENCE: AL-2-C3  
; CURRENT APPLICATION NUMBER: US/09/292,225  
; CURRENT FILING DATE: 1999-04-15  
; EARLIER APPLICATION NUMBER: 60/098,909  
; EARLIER FILING DATE: 1998-09-02  
; EARLIER APPLICATION NUMBER: 60/085,295  
; EARLIER FILING DATE: 1998-05-13  
; EARLIER APPLICATION NUMBER: 60/098,565  
; EARLIER FILING DATE: 1998-04-17  
; EARLIER APPLICATION NUMBER: 09/062,013  
; EARLIER FILING DATE: 1998-04-17  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 15  
; LENGTH: 555  
; TYPE: PRT  
; ORGANISM: Dermatophagoides farinae  
US-09-292-225-15

Query Match 100.0%; Score 78; DB 4; Length 555;  
Best Local Similarity 100.0%; Pred. No. 9.6e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIKRDNDYSKNPM 14  
DB 20 SIKRDNDYSKNPM 33

RESULT 5  
US-09-292-225-18

; Sequence 18, Application US/09292225  
; Patent No. 6455686  
; GENERAL INFORMATION:  
; APPLICANT: McCall, Catherine A.  
; APPLICANT: Hunter, Shirley Wu  
; APPLICANT: Weber, Eric R.  
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
; FILE REFERENCE: AL-2-C3  
; CURRENT APPLICATION NUMBER: US/09/292,225  
; CURRENT FILING DATE: 1999-04-15  
; EARLIER APPLICATION NUMBER: 60/098,909  
; EARLIER FILING DATE: 1998-09-02  
; EARLIER APPLICATION NUMBER: 60/085,295  
; EARLIER FILING DATE: 1998-05-13  
; EARLIER APPLICATION NUMBER: 60/098,565  
; EARLIER FILING DATE: 1998-04-17  
; EARLIER APPLICATION NUMBER: 09/062,013  
; EARLIER FILING DATE: 1998-04-17  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 18  
; LENGTH: 555  
; TYPE: PRT  
; ORGANISM: Dermatophagoides farinae  
US-09-292-225-18

Query Match 100.0%; Score 78; DB 4; Length 555;  
Best Local Similarity 100.0%; Pred. No. 9.6e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SIKRDNDYSKNPM 14  
DB 20 SIKRDNDYSKNPM 33

RESULT 6  
US-09-292-225-41  
; Sequence 41, Application US/09292225



```
/ Patent No. 6455686
/ GENERAL INFORMATION:
/ APPLICANT: McCall, Catherine A.
/ APPLICANT: Hunter, Shirley Wu
/ APPLICANT: Weber, Eric R.
/ TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
/ FILE REFERENCE: AL-2-C3
/ CURRENT APPLICATION NUMBER: US/09/292,225
/ EARLIER FILING DATE: 1999-04-15
/ EARLIER APPLICATION NUMBER: 60/098,909
/ EARLIER FILING DATE: 1998-09-02
/ EARLIER APPLICATION NUMBER: 60/085,295
/ EARLIER FILING DATE: 1998-05-13
/ EARLIER APPLICATION NUMBER: 60/098,565
/ EARLIER FILING DATE: 1998-04-17
/ EARLIER APPLICATION NUMBER: 09/062,013
/ EARLIER FILING DATE: 1998-04-17
/ NUMBER OF SEQ ID NOS: 49
/ SOFTWARE: Patentln Ver. 2.0
/ SEQ ID NO 41
/ LENGTH: 490
/ TYPE: PRT
/ ORGANISM: Dermatophagoides farinae
US-09-292-225-41
```

```
Query Match      83.3%; Score 65; DB 4; Length 490;
Best Local Similarity 91.7%; Pred. No. 0.0017;
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      3 KRDNHNSKNPM 14
DB      3 KRDNHNSKNPM 14

RESULT 7
US-09-292-225-35
/ Sequence 35; Application US/09292225
/ Patent No. 6455686
/ GENERAL INFORMATION:
/ APPLICANT: McCall, Catherine A.
/ APPLICANT: Hunter, Shirley Wu
/ APPLICANT: Weber, Eric R.
/ TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
/ FILE REFERENCE: AL-2-C3
/ CURRENT APPLICATION NUMBER: US/09/292,225
/ EARLIER FILING DATE: 1999-04-15
/ EARLIER APPLICATION NUMBER: 60/098,909
/ EARLIER FILING DATE: 1998-09-02
/ EARLIER APPLICATION NUMBER: 60/085,295
/ EARLIER FILING DATE: 1998-05-13
/ EARLIER APPLICATION NUMBER: 60/098,565
/ EARLIER FILING DATE: 1998-04-17
/ EARLIER APPLICATION NUMBER: 09/062,013
/ EARLIER FILING DATE: 1998-04-17
/ NUMBER OF SEQ ID NOS: 49
/ SOFTWARE: Patentln Ver. 2.0
/ SEQ ID NO 35
/ LENGTH: 509
/ TYPE: PRT
/ ORGANISM: Dermatophagoides farinae
US-09-292-225-35
```

```
Query Match      83.3%; Score 65; DB 4; Length 509;
Best Local Similarity 91.7%; Pred. No. 0.0017;
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      3 KRDNHNSKNPM 14
DB      22 KRDNHNSKNPM 33
```

```
RESULT 8
US-09-292-225-38
/ Sequence 38; Application US/09292225
/ Patent No. 6455686
/ GENERAL INFORMATION:
/ APPLICANT: McCall, Catherine A.
/ APPLICANT: Hunter, Shirley Wu
/ APPLICANT: Weber, Eric R.
/ TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
/ FILE REFERENCE: AL-2-C3
/ CURRENT APPLICATION NUMBER: US/09/292,225
/ EARLIER FILING DATE: 1999-04-15
/ EARLIER APPLICATION NUMBER: 60/098,909
/ EARLIER FILING DATE: 1998-09-02
/ EARLIER APPLICATION NUMBER: 60/085,295
/ EARLIER FILING DATE: 1998-05-13
/ EARLIER APPLICATION NUMBER: 60/098,565
/ EARLIER FILING DATE: 1998-04-17
/ EARLIER APPLICATION NUMBER: 09/062,013
/ EARLIER FILING DATE: 1998-04-17
/ NUMBER OF SEQ ID NOS: 49
/ SOFTWARE: Patentln Ver. 2.0
/ SEQ ID NO 38
/ LENGTH: 509
/ TYPE: PRT
/ ORGANISM: Dermatophagoides farinae
US-09-292-225-38
```

```
Query Match      83.3%; Score 65; DB 4; Length 509;
Best Local Similarity 91.7%; Pred. No. 0.0017;
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      3 KRDNHNSKNPM 14
DB      22 KRDNHNSKNPM 33

RESULT 9
US-09-543-681A-7570
/ Sequence 7570; Application US/09543681A
/ Patent No. 6605709
/ GENERAL INFORMATION:
/ APPLICANT: GARY BRETON
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
/ FILE REFERENCE: 2709,1002-001
/ CURRENT APPLICATION NUMBER: US/09/543,681A
/ EARLIER FILING DATE: 2000-04-05
/ PRIOR APPLICATION NUMBER: US 60/128,706
/ EARLIER FILING DATE: 1999-04-09
/ NUMBER OF SEQ ID NOS: 8344
/ SEQ ID NO 7570
/ LENGTH: 121
/ TYPE: PRT
/ ORGANISM: Proteus mirabilis
US-09-543-681A-7570
```

```
Query Match      52.6%; Score 41; DB 4; Length 121;
Best Local Similarity 58.3%; Pred. No. 5.9;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 SIKRDNHNSKNV 12
DB      15 SYKRDHNSKNPM 26
```

```
RESULT 10
US-09-252-991A-25153
/ Sequence 25153; Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
```

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 25153  
LENGTH: 820  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-25153

Query Match 50.0%; Score 39; DB 4; Length 820;  
Best Local Similarity 46.2%; Pred. No. 1.1e+02;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 IKRDNDYSKNP 14  
DB 372 VRVHSSYSARPM 384

RESULT 11  
US-09-134-001C-5099  
Sequence 5099, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 5099  
LENGTH: 150  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5099

Query Match 48.7%; Score 38; DB 4; Length 150;  
Best Local Similarity 41.7%; Pred. No. 25;  
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 IKRDNDYSKNP 13  
DB 40 LKQENKXKXENP 51

RESULT 12  
US-09-252-991A-24219  
Sequence 24219, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 24219

LENGTH: 203  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-24219

Query Match 48.7%; Score 38; DB 4; Length 203;  
Best Local Similarity 46.2%; Pred. No. 35;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 IKRDNDYSKNP 14  
DB 98 LARDNDYADEPL 110

RESULT 13  
US-09-371-913A-7  
Sequence 7, Application US/09371913A  
Patent No. 629369  
GENERAL INFORMATION:  
APPLICANT: Schnepf, H. Ernest  
APPLICANT: Narva, Kenneth E.  
APPLICANT: Stockhoff, Brian A.  
APPLICANT: Pinstad Lee, Stacey  
APPLICANT: Walz, Mikki  
APPLICANT: Sturgis, Blake  
TITLE OF INVENTION: Pesticidal Toxins and Genes from Bacillus laterosporus  
FILE REFERENCE: MA-719XC2  
CURRENT APPLICATION NUMBER: US/09/371,913A  
CURRENT FILING DATE: 1999-08-10  
PRIOR APPLICATION NUMBER: 60/095,955  
PRIOR FILING DATE: 1998-08-10  
PRIOR APPLICATION NUMBER: 60/138,251  
PRIOR FILING DATE: 1999-06-08  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 7  
LENGTH: 784  
TYPE: PRT  
ORGANISM: Peptide sequence  
US-09-371-913A-7

Query Match 48.7%; Score 38; DB 3; Length 784;  
Best Local Similarity 66.7%; Pred. No. 1.6e+02;  
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 RDNDYSKN 12  
DB 758 QDVNYSKN 766

RESULT 14  
US-09-967-805-7  
Sequence 7, Application US/09967805  
Patent No. 6605701  
GENERAL INFORMATION:  
APPLICANT: Schnepf, Ernest H  
APPLICANT: Narva, Kenneth E.  
APPLICANT: Stockhoff, Brian A.  
APPLICANT: Pinstad Lee, Stacey  
APPLICANT: Walz, Mikki  
APPLICANT: Sturgis, Blake  
TITLE OF INVENTION: Pesticidal Toxins and Genes from Bacillus laterosporus  
FILE REFERENCE: MA-719XC2D1  
CURRENT APPLICATION NUMBER: US/09/967,805  
CURRENT FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: 09/371,913  
PRIOR FILING DATE: 1999-08-10  
PRIOR APPLICATION NUMBER: 60/095,955  
PRIOR FILING DATE: 1998-08-10  
PRIOR APPLICATION NUMBER: 60/138,251  
PRIOR FILING DATE: 1999-06-08

; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO 7  
 ; LENGTH: 784  
 ; TYPE: PRT  
 ; ORGANISM: Peptide sequence  
 US-09-967-805-7

Query Match 48.7%; Score 38; DB 4; Length 784;  
 Best Local Similarity 66.7%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 RDHNDYSKN 12  
 :|:|:|:|  
 Db 758 QDYNYSKN 766

RESULT 15  
 US-09-719-190-2  
 ; Sequence 2, Application US/09719190  
 ; Patent No. 6649171  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Thonard, Joelle  
 ; TITLE OF INVENTION: Moraxella Catharrahalis Polynucleotides  
 ; TITLE OF INVENTION: and Polypeptides  
 ; FILE REFERENCE: BM45326  
 ; CURRENT APPLICATION NUMBER: US/09/719,190  
 ; PRIOR FILING DATE: 2001-05-14  
 ; PRIOR APPLICATION NUMBER: PCT/EP99/03824  
 ; PRIOR FILING DATE: 1999-05-31  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 947  
 ; TYPE: PRT  
 ; ORGANISM: Moraxella catharrahalis  
 US-09-719-190-2

Query Match 48.7%; Score 38; DB 4; Length 947;  
 Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 HNDYSKN 12  
 ||:|:|:|  
 Db 468 HNEYSKN 474

Search completed: March 22, 2004, 07:03:52  
 Job time : 1.1284 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:40:28 ; Search time 1.30072 Seconds  
(without alignments)

1479.047 Million cell updates/sec

Title: US-09-662-293-3

Perfect score: 115  
Sequence: 1 DIPPTNIHKYLVCESYNGG 20

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	48.7	1635	T14075	chitinase (EC 3.2.
2	53.5	46.5	66	T13181	exiclionase - Lacto
3	53	46.1	525	T14445	chitinase (EC 3.2.
4	47	40.9	291	AC1136	shikimate 5-dehydr
5	47	40.9	291	AS1494	DNA methyltransfer
6	47	40.9	330	S5390	glycoprotein H - h
7	47	40.9	841	VGEB37	cdp synthase (pyrg
8	46	40.0	586	C71695	protein kinase kin
9	46	40.0	891	T140503	protein kinase 1 -
10	46	40.0	891	A28903	hypothetical prote
11	46	39.6	995	T05842	hypothetical prote
12	45.5	39.6	342	T40090	ACMNPV orf150 - Bo
13	45	39.1	115	T14187	chitinase (EC 3.2.
14	45	39.1	504	A28221	probable ubiquitin
15	45	39.1	523	G84732	transcription regu
16	45	39.1	867	S57795	early B1A 20X prot
17	44	38.3	91	D97130	venombin A (EC 3.4
18	44	38.3	171	B60010	lag-2 proteain - Ca
19	44	38.3	231	A60468	sugar phosphatase - fi
20	44	38.3	402	S42367	citrate lyase - fi
21	44	38.3	407	A29223	insecticidal toxin
22	44	38.3	492	T38156	probable ribosomal
23	44	38.3	695	D86392	hypothetical prote
24	43.5	37.8	1496	AH0447	hypothetical prote
25	43	37.4	102	F71354	protein kinase (EC
26	43	37.4	304	T23919	
27	43	37.4	495	S60761	
28	43	37.4	502	D86765	
29	43	37.4	683	A23690	

30	43	37.4	683	1	S24478	protein kinase C (
31	43	37.4	897	2	H86905	cation-transportin
32	42.5	37.0	533	2	G69735	xylian 1,4-beta-xy
33	42	36.5	76	2	AH2310	hypothetical prote
34	42	36.5	99	2	H72868	AcOrf-150 protein
35	42	36.5	116	2	UC7620	guanylin precursor
36	42	36.5	202	2	C88989	protein F02C9.4 (i
37	42	36.5	236	1	A32121	snake venom factor
38	42	36.5	236	1	B32121	snake venom factor
39	42	36.5	254	2	T18987	hypothetical prote
40	42	36.5	257	1	UC2479	venombin B (EC 3.4
41	42	36.5	317	2	S00019	L-lactate dehydrog
42	42	36.5	401	2	T24929	hypothetical prote
43	42	36.5	462	2	A72521	probable glutamine
44	42	36.5	743	2	A97021	pyruvate-formate 1
45	42	36.5	856	2	T43631	serine/threonine k

#### ALIGNMENTS

##### RESULT 1

T14075

Chitinase (EC 3.2.1.14) - yellow fever mosquito

C/Species: Aedes aegypti (yellow fever mosquito)

C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C/Accession: T14075

R/de la Vega, H.; Specht, C.A.; Liu, Y.; Robbins, P.W.

Insect Mol. Biol. 7, 233-239, 1997

A/Title: Chitinases are a multi-gene family in Aedes, Anopheles, and Drosophila.

A/Reference number: Z17872

A/Accession: T14075

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1635 <DEU>

A/Cross-references: EMBL:AF026492; NID:G2564720; PID:G2564721; PIDN:AAB81850.1

A/Gene: CRT2

A/Intons: 462/3; 524/3; 618/1; 951/3; 1151/2

C/Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 48.7%; Score 56; DB 2; Length 1635;

Best Local Similarity 50.0%; Pred. No. 3;

Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 IPHPTNIHKYLVCE 15

DB 1134 VPHPTDKRYTICQ 1147

##### RESULT 2

T13181

exiclionase - Lactobacillus phage phi-gle

C/Species: Lactobacillus phage phi-gle

C/Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 13-Aug-1999

C/Accession: T13181

R/Kodaira, K.I.; Oki, M.; Kakikawa, M.; Watanabe, N.; Hiraakawa, M.; Yamada, K.; Takeo, Gene 187, 45-53, 1997

A/Title: Genome structure of the Lactobacillus temperate phage phi gle: the whole genom

A/Reference number: Z17631; WUID:97225795; PMID:9073065

A/Accession: T13181

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 166 <KOD>

A/Cross-references: EMBL:X58106; NID:e917136; PID:e247183; PIDN:CAA66757.1

A/Gene: Xis

Query Match 46.5%; Score 53.5; DB 2; Length 66;

Best Local Similarity 50.0%; Pred. No. 0.28;

Matches 9; Conservative 4; Mismatches 2; Indels 3; Gaps 1;

QY 1 DIPPTNI--HKYLVCE 15

Db 37 DVFPNTINASTYRFLACD 54

### RESULT 3

T44445  
chitinase (EC 3.2.1.14) [imported] - African malaria mosquito  
C/Species: Anopheles gambiae (African malaria mosquito)  
C/Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jan-2000  
C/Accession: T44445  
R:Shen, Z.; Jacobs-Lorena, M.  
submitted to the EMBL Data Library, June 1997  
A/Reference number: Z22771  
A/Accession: T44445  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-525 <SHE>  
A/Cross-references: EMBL:AF008575; PDB:1AB87764.1  
A/Experimental source: adult; gut  
C/Genetics:  
A/Gene: chi-1  
C/Keywords: glycosidase; hydrolase

Query Match 46.1%; Score 53; DB 2; Length 525;  
Best Local Similarity 53.8%; Pred. No. 2.8;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 IPHPTNHHKYL 14  
DB 478 VPHPTNHCARYIC 490

### RESULT 4

shikimate 5-dehydrogenase homolog lmo0490 [imported] - Listeria monocytogenes (strain EG  
C/Species: Listeria monocytogenes  
C/Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 07-Jul-2003  
C/Accession: AC1136  
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A/Authors: Krefelt, T.; Kuhn, M.; Kunst, F.; Kurapkai, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voess, H.; Wehlend,  
A./Title: Comparative genomics of Listeria species.  
A/Reference number: AB1077; MUID:21537279; PMID:11679669  
A/Accession: AC1136  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-291 <GUA>  
A/Cross-references: GB:NC\_003210; PDB:1G16409866; GSPDB:GN00177  
A/Experimental source: strain EGD-e  
C/Genetics:  
A/Gene: lmo0490  
C/Superfamily: shikimate 5-dehydrogenase; shikimate dehydrogenase homology

Query Match 40.9%; Score 47; DB 2; Length 291;  
Best Local Similarity 72.7%; Pred. No. 13;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 IPHPTNHHKYL 12  
DB 74 MPNKTNIHXYL 84

### RESULT 5

AE1494  
shikimate 5-dehydrogenase homolog lmo0493 [imported] - Listeria innocua (strain Clip1126  
C/Species: Listeria innocua  
C/Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 07-Jul-2003  
C/Accession: AE1494  
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.

D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A/Authors: Krefelt, T.; Kuhn, M.; Kunst, F.; Kurapkai, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voess, H.; Wehlend,  
A./Title: Comparative genomics of Listeria species.  
A/Reference number: AB1077; MUID:21537279; PMID:11679669  
A/Accession: AE1494  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-291 <GUA>  
A/Cross-references: GB:AL592022; PDB:1G16412933; GSPDB:GN00178  
A/Experimental source: strain Clip11262  
C/Genetics:  
A/Gene: lmo0493  
C/Superfamily: shikimate 5-dehydrogenase; shikimate dehydrogenase homology

Query Match 40.9%; Score 47; DB 2; Length 291;  
Best Local Similarity 72.7%; Pred. No. 13;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 IPHPTNHHKYL 12  
DB 74 MPNKTNIHXYL 84

### RESULT 6

S53990  
DNA methyltransferase pmt1 - fission yeast (Schizosaccharomyces pombe)  
C/Species: Schizosaccharomyces pombe  
C/Date: 15-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 17-Mar-2000  
C/Accession: S53990; T39793  
R:Wilkinson, C.R.M.; Bartlett, R.; Nurse, P.; Bird, A.P.  
Nucleic Acids Res. 23, 203-210, 1995  
A/Title: The fission yeast gene pmt1(+) encodes a DNA methyltransferase homologue.  
A/Reference number: S53990; MUID:5516638; PMID:7662522  
A/Accession: S53990  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-330 <WLS>  
A/Cross-references: EMBL:X82444; NID:G563910; PDB:1CA57824.1; PID:G563911  
R:McDougal, R.C.; Rajandream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.  
submitted to the EMBL Data Library, July 1999  
A/Reference number: Z21880  
A/Accession: T39793  
A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA  
A/Residues: 1-330 <MCD>  
A/Cross-references: EMBL:AL109731; PDB:1CA57824.1; GSPDB:GN00067; SPDB:SPBC19C2.02  
A/Experimental source: strain 972h-; cosmid c19C2  
C/Genetics:  
A/Gene: SPBC19C2.02  
A/Map position: 2  
A/Intons: 56/3  
C/Superfamily: site-specific methyltransferase (cytosine-specific) EcoRII

Query Match 40.9%; Score 47; DB 2; Length 330;  
Best Local Similarity 38.9%; Pred. No. 15;  
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 2 IPHPTNHHKYL 19  
DB 108 LPHVNNLPYILINVOG 125

### RESULT 7

VGEB37  
glycoprotein H - human herpesvirus 3  
N/Alternate names: glycoprotein III  
C/Species: human herpesvirus 3, varicella-zoster virus  
C/Date: 30-Sep-1988 #sequence\_revision 30-Sep-1988 #text\_change 16-Jul-1999  
C/Accession: B27341  
R:Davidson, A.J.; Scott, J.E.  
J. Gen. Virol. 67, 1759-1816, 1986

A>Title: The complete DNA sequence of varicella-zoster virus.  
 A/Reference number: A27345; MUID:86306657; PMID:3018124  
 A/Accession: B27341  
 A/Molecule type: DNA  
 A/Residues: 1-841 <DAV>  
 A/Cross-references: EMBL:X04370; NID:G59989; PIDN:CAA27920.1; PID:G60026  
 C/Genetics:  
 A/Gene: 37  
 C/Superfamily: herpesvirus glycoprotein H  
 C/Species: Schizosaccharomyces pombe  
 C/Keywords: glycoprotein  
 F,18,45,83,217,317,499,522,560,760,783/Binding site: carbohydrate (asn) (covalent) #stat

Query Match 40.9%; Score 47; DB 1; Length 841;  
 Best Local Similarity 50.0%; Pred. No. 38;  
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 2 IPHPTNKHLYVCEV 17  
 DB 715 LPHDNLKELVCGSV 730

RESULT 8  
 C71695  
 ctp synthase (pyrg) RP378 - Rickettsia prowazekii  
 C/Species: Rickettsia prowazekii  
 C/Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000  
 C/Accession: C71695  
 R/Anderson, S.G.E.; Zomrodipour, A.; Anderson, J.O.; Sichevitz-Ponten, T.; Almark, U  
 Nature 396, 133-140, 1998  
 A/Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.  
 A/Reference number: A71630; MUID:99039499; PMID:9823893  
 A/Accession: C71695  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-586 <AND>  
 A/Cross-references: GB:AJ235271; GB:AJ235269; NID:G3868717; PIDN:CAA14837.1; PID:G386093  
 A/Experimental source: strain Madrid E  
 C/Genetics:  
 A/Gene: PYrg, RP378  
 C/Superfamily: CTP synthase

Query Match 40.0%; Score 46; DB 2; Length 586;  
 Best Local Similarity 50.0%; Pred. No. 38;  
 Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 2 IPHPTNKHLYVCEV 19  
 DB 115 IPHVTNLIKDFIMSTNG 132

RESULT 9  
 T40503  
 protein kinase kin - fission yeast (Schizosaccharomyces pombe)  
 C/Species: Schizosaccharomyces pombe  
 C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Jan-2000  
 C/Accession: T40503  
 R/Gwilliam, R.; Rajadream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.  
 submitted to the EMBL Data Library, September 1998  
 A/Reference number: Z21933  
 A/Accession: T40503  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-891 <GMT>  
 A/Cross-references: EMBL:AL031534; PIDN:CAA20726.1; GSPDB:GN00067; SPDB:SPBC4F6.06  
 A/Experimental source: strain 972h-; cosmid c4F6  
 C/Genetics:  
 A/Gene: SPDB:SPBC4F6.06  
 A/Map position: 2  
 C/Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo

Query Match 40.0%; Score 46; DB 2; Length 891;  
 Best Local Similarity 60.0%; Pred. No. 58;  
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 6 TNIHKYLVCEVNG 20  
 DB 212 TNSHYVWFVVDG 226

RESULT 10  
 A38903  
 protein kinase 1 - fission yeast (Schizosaccharomyces pombe)  
 C/Species: Schizosaccharomyces pombe  
 C/Date: 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change 24-Sep-1999  
 C/Accession: A38903; A36474  
 R/Levin, D.E.; Bishop, J.M.  
 submitted to Genbank, May 1991  
 A/Reference number: A38903  
 A/Accession: A38903  
 A/Molecule type: DNA  
 A/Residues: 1-891 <LEV>  
 A/Cross-references: GB:M64999; NID:G173409; PIDN:AA63577.1; PID:G173410  
 R/Levin, D.E.; Bishop, J.M.  
 Proc. Natl. Acad. Sci. U.S.A. 87, 8272-8276, 1990  
 A/Title: A putative protein kinase gene (kin1(+)) is important for growth polarity in  
 A/Reference number: A36474; MUID:91045979; PMID:2236039  
 A/Accession: A36474  
 A/Status: nucleic acid sequence not shown  
 A/Molecule type: DNA  
 A/Residues: 1-361, 'R', 363-619, 'I', 621-707, 'W', 709-891 <LEV>  
 A/Cross-references: GB:M36060  
 C/Genetics:  
 A/Gene: kin1+  
 C/Superfamily: unassigned ser/thr or tyr-specific protein kinases; protein kinase homol  
 C/Keywords: ATP; phosphotransferase; serine/threonine-specific protein kinase  
 F,113-195/Domain: protein kinase ATP-binding motif  
 F,111-139/Region: protein kinase ATP-binding motif

Query Match 40.0%; Score 46; DB 2; Length 891;  
 Best Local Similarity 60.0%; Pred. No. 58;  
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 6 TNIHKYLVCEVNG 20  
 DB 212 TNSHYVWFVVDG 226

RESULT 11  
 T05842  
 hypothetical protein F17L22.130 - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 09-Apr-1999 #sequence\_revision 09-Apr-1999 #text\_change 13-Aug-1999  
 C/Accession: T05842  
 R/Bevan, M.; Van Der Schueren, J.; Chuang, Y.U.; Voet, M.; Robben, J.; Volckaert, G.; E  
 submitted to the Protein Sequence Database, February 1999  
 A/Reference number: Z15454  
 A/Accession: T05842  
 A/Molecule type: DNA  
 A/Residues: 1-995 <BEV>  
 A/Cross-references: EMBL:AL035527  
 A/Experimental source: cultivar Columbia; BAC clone F17L22  
 C/Genetics:  
 A/Map position: 4  
 A/Intons: 106/3; 265/3; 350/1; 402/3; 425/2; 460/3; 488/3; 690/3; 712/1; 755/3; 773/3;  
 A/Note: F17L22.130

Query Match 40.0%; Score 46; DB 2; Length 995;  
 Best Local Similarity 45.0%; Pred. No. 65;  
 Matches 9; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

OY 1 DIPPTNKHLYVCEVNG 20  
 DB 446 DIPSPDVSHYVSEDDTSG 465

RESULT 12

T40090  
 Hypothetical protein SPBC29A3.19 - fission yeast (*Schizosaccharomyces pombe*)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 20-Oct-2000  
 C:Accession: T40090; T39752  
 R:Llyne, M.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.  
 submitted to the EMBL Data Library, March 1998  
 A:Reference number: 221904  
 A:Accession: T40090  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-342 <LYN>  
 A:Cross-references: EMBL:AL022299; PIDN:CAA18396.1; GSPDB:GN00067; SPDB:SPBC29A3.19  
 A:Experimental source: strain 972h-; cosmid c29A3  
 R:Wood, V.; Barrell, B.G.; Rajandream, M.A.; Xiang, Z.; Aves, S.  
 submitted to the EMBL Data Library, January 1999  
 A:Reference number: 221878  
 A:Accession: T39752  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 198-342 <WOC>  
 A:Cross-references: EMBL:AL035077; PIDN:CAA22661.1; GSPDB:GN00066; SPDB:SPBC18E5.01  
 A:Experimental source: strain 972h-; cosmid c18E5  
 C:Genetics: <LYN1>  
 A:Gene: SPDB:SPBC29A3.19  
 A:Map position: 2  
 C:Genetics: <WOC1>  
 A:Gene: SPDB:SPBC18E5.01  
 A:Map position: 1

Query Match 39.6%; Score 45.5; DB 2; Length 342;  
 Best Local Similarity 55.6%; Pred. No. 26;  
 Matches 10; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 3 PHEPTNHHKYLVCESVNG 20  
 DB 130 PHQVTAHKLVC-SVDLG 146

RESULT 13  
 T41887  
 ACNMPV.0r150 - Bombyx mori nuclear polyhedrosis virus (isolate T3)  
 C:Species: Bombyx mori nuclear polyhedrosis virus; BMSNPV  
 A:Variety: isolate T3  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 11-May-2000  
 C:Accession: T41887  
 R:Gomi, S.; Majima, K.; Maeda, S.  
 J. Gen. Virol. 80, 1323-1337, 1999  
 A:Title: Sequence analysis of the genome of Bombyx mori nucleopolyhedrovirus.  
 A:Reference number: Z22020; MUID:99281911; PMID:10355780  
 A:Accession: T41887  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-115 <RAM>  
 A:Cross-references: EMBL:U33180; PIDN:AAC63816.1  
 A:Experimental source: isolate T3  
 C:Genetics:  
 A:Note: Orf\_126

Query Match 39.1%; Score 45; DB 2; Length 115;  
 Best Local Similarity 37.5%; Pred. No. 10;  
 Matches 6; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 3 PHEPTNHHKYLVCESVN 18  
 DB 60 PHPTKCNAPYCWGIN 75

RESULT 14  
 A38221  
 chitinase (EC 3.2.1.14) MFL - nematode (*Brugia malayi*)  
 C:Species: Brugia malayi  
 C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 21-Jul-2000

C:Accession: A38221  
 R:Fullman, J.A.; Lane, W.S.; Smith, R.F.; Piessens, W.F.; Perler, F.B.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 1548-1552, 1992  
 A:Title: Transmembrane blocking antibodies recognize microfilarial chitinase in brugian I  
 A:Reference number: A38221; MUID:92179220; PMID:1542646  
 A:Accession: A38221  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid; protein  
 A:Residues: 1-504 <FHM>  
 A:Cross-references: GB:M73689; NID:G156063; PIDN:AAA27854.1; PID:G156064  
 A:Note: sequence extracted from NCBI backbone (NCBI:85345)  
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 39.1%; Score 45; DB 2; Length 504;  
 Best Local Similarity 50.0%; Pred. No. 46;  
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 PHEPTNHHKYLVC 14  
 DB 459 PHPTDCHLFRQC 470

RESULT 15  
 G84732  
 probable ubiquitin activating enzyme [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: G84732  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.T.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84420; MUID:20083487; PMID:10617197  
 A:Accession: G84732  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-523 <STO>  
 A:Cross-references: GB:AE002093; NID:G3831455; PIDN:AAC69937.1; GSPDB:GN00139  
 A:Genetics:  
 A:Gene: Atg32410  
 A:Map position: 2

Query Match 39.1%; Score 45; DB 2; Length 523;  
 Best Local Similarity 47.1%; Pred. No. 48;  
 Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 PHEPTNHHKYLVCESVNG 19  
 DB 393 PSTTEHKLADENYSG 409

Search completed: March 22, 2004, 07:01:19  
 Job time : 3.30072 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:31:13 ; Search time 0.767093 Seconds  
(Without alignments)  
1357.597 Million cell updates/sec

Title: US-09-662-293-3  
Perfect score: 115  
Sequence: 1 DIHPNTNKHVCESYNGG 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	47	40.9	258 1	VSP3_BOTIA
2	47	40.9	291 1	AROE_LISTIN
3	47	40.9	291 1	AROE_LISTMO
4	47	40.9	330 1	PMT1_SCHPO
5	47	40.9	841 1	VGLH_VZVD
6	46	40.0	153 1	PSI_ANOCH
7	46	40.0	586 1	PRG_RICPR
8	46	40.0	891 1	KINI_SCHPO
9	46	40.0	1934 1	PNO_CRYPV
10	45.5	39.6	342 1	YBSI_SCHPO
11	45	39.1	458 1	CHIT3_DROME
12	45	39.1	504 1	CHIT_BROMA
13	44	38.3	103 1	RS10_FUSBN
14	44	38.3	171 1	EIA_ADECC
15	44	38.3	231 1	VSP1_AGKCO
16	44	38.3	402 1	LAG2_CAEEL
17	44	38.3	624 1	SVV4_ARATH
18	44	38.3	772 1	KEA4_HUMAN
19	43	37.4	102 1	RS10_TREBA
20	43	37.4	258 1	VSP3_TRIGA
21	43	37.4	434 1	BRAC_HEMPU
22	43	37.4	683 1	KPCL_MOUSE
23	43	37.0	683 1	KPCL_RAT
24	42.5	37.0	935 1	AD22_XENLA
25	42	36.5	99 1	Y150_NPVAC
26	42	36.5	236 1	VSP4_DABRU
27	42	36.5	236 1	VSPG_DABRU
28	42	36.5	252 1	TPIS_STRA3
29	42	36.5	252 1	TPIS_STRA3
30	42	36.5	252 1	TPIS_STRA3
31	42	36.5	257 1	VSP1_TRIMU
32	42	36.5	257 1	VSP3_TRIMU
33	42	36.5	260 1	VSP2_VIPLE

34	42	36.5	317 1	LDH_BACCA	P10655 bacillus ca
35	42	36.5	347 1	DHAS_LEGPN	O31219 legioneila
36	42	36.5	518 1	ASB3_HUMAN	O9y575 homo sapien
37	42	36.5	736 1	STB3_HUMAN	O81wq3 homo sapien
38	42	36.5	794 1	KI11_HUMAN	O8tdc3 homo sapien
39	42	36.5	856 1	UN51_CAEEL	O23023 caenorhabdi
40	42	36.5	3066 1	POLG_SBMVN	O90069 s genome po
41	42	36.5	434 1	NADB_CLOAB	P21231 s genome po
42	41.5	36.1	581 1	FRZ3_DROME	O97k95 clostridium
43	41.5	36.1	581 1	FRZ3_DROME	O77438 drosophila
44	41.5	36.1	703 1	YK76_CAEEL	P34317 caenorhabdi
45	41	35.7	103 1	RS10_BORBU	P94266 borrelia bu

ALIGNMENTS

RESULT 1	VSP3_BOTIA	STANDARD;	PRT;	258 AA.
ID	VSP3_BOTIA			
AC	Q9PT08;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Venom serine proteinase A precursor (EC 3.4.21.-).			
OS	Bothrops jararaca (Viperidae).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;			
OC	Viperidae; Crotalinae; Bothrops.			
OX	NCBI_TaxID=8724;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Venom gland;			
RA	Murayama N.;			
RL	Submitted (Aug-1999) to the EMBL/Genbank/DBJ databases.			
CC	- FUNCTION: Thrombin-like snake venom serine protease.			
CC	- SUBCELLULAR LOCATION: Secreted.			
CC	- TISSUE SPECIFICITY: Expressed by the venom gland.			
CC	- SIMILARITY: Belongs to peptidase family S1. Snake venom subfamily.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).			
CC	-----			
DR	EMBL; AB031394; BAAB9310.1; -.			
DR	HSSP; P00763; IDPO.			
DR	MEROPS; S01.181; -.			
DR	InterPro; IPR009003; Cys_Ser_trypsin.			
DR	InterPro; IPR001254; Peptidase_S1.			
DR	InterPro; IPR001314; Peptidase_S1A.			
DR	Pfam; PF00089; trypsin; 1.			
DR	PRINTS; PR00722; CHYOTRYPsin.			
DR	SMART; SM00020; TRYP_SPC; 1.			
DR	PROSITE; PS50240; TRYPsin_DOM; 1.			
DR	PROSITE; PS00134; TRYPsin_HIS; 1.			
DR	PROSITE; PS00135; TRYPsin_SER; 1.			
KW	Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal.			
FT	STGMN 1			
FT	PROPER 19			
FT	CHAIN 25			
FT	ACT_SITE 65			
FT	ACT_SITE 110			
FT	ACT_SITE 204			
FT	DISULFID 31			
FT	DISULFID 50			
FT	DISULFID 66			
FT	DISULFID 98			
FT	DISULFID 142			
FT	DISULFID 174			
FT	DISULFID 200			
FT	DISULFID 225			



```

FT CARBOHYD 32 32 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 44 44 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 154 154 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 170 170 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 251 251 N-LINKED (GLCNAC. . .) (POTENTIAL)
SQ SEQUENCE 258 AA; 28058 MW; 034F49PDCB79EB64 CRC64;

```

```

Query Match 40.9%; Score 47; DB 1; Length 258;
Best Local Similarity 40.9%; Pred. No. 3.2;
Matches 9; Conservative 3; Mismatches 8; Indels 2; Gaps 1;

```

```

Qy 1 DIPHTNH-KYLVCSYNG 20
Db 159 DVPHCNILFMYTCRGAHAG 180

```

## RESULT 2

```

ID AROE LISIN STANDARD; PRT; 291 AA.
AC 092RG7;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Shikimate 5-dehydrogenase (EC 1.1.1.25).
GN AROE OR LMO0493.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
NCBI_TaxID=1642;
RX SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 6a;
RX MEDLINE=21537279; PubMed=11679669;
RA Baquero F., Franquet L., Buchrieser C., Rusniok C., Amend A.,
RA Glasier P., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Darvar A., Deboux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussauget O.,
RA Entian K.-D., Fahl H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf U., Jackson D.,
RA Jones L.-M., Kaerst U., Krell T., Kuhn M., Kunst F., Kurapkak G.,
RA Madueno E., Maitounam A., Mata Vicente U., Ng E., Negtari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
CC -1- CATALYTIC ACTIVITY: Shikimate + NADP(+) = 5-dehydroshikimate +
CC NADPH.
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC -1- fourth step.
CC -1- SIMILARITY: Belongs to the shikimate dehydrogenase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL596165; CAC95725.1; -.
CC PIR; A61494; A61494.
CC Listlist; LMO0493; -.
CC HAMAP; MF_00222; -; 1.
CC InterPro; IPR006151; Shikimate_DH.
CC Pfam; PF01488; Shikimate_DH; 1.
CC KMW Aromatic amino acid biosynthesis; Oxidoreductase; NADP;
CC Complete proteome.
SQ SEQUENCE 291 AA; 32192 MW; 119148B1DA3F244 CRC64;

```

```

Query Match 40.9%; Score 47; DB 1; Length 291;

```

```

Best Local Similarity 72.7%; Pred. No. 3.6;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 2 DIPHTNHKYL 12
Db 74 MPKNTNHHKYL 84

```

## RESULT 3

```

ID AROE LISMO STANDARD; PRT; 291 AA.
AC 08Y9N5;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Shikimate 5-dehydrogenase (EC 1.1.1.25).
GN AROE OR LMO0490.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
NCBI_TaxID=1639;
RX SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Baquero F., Franquet L., Buchrieser C., Rusniok C., Amend A.,
RA Glasier P., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Darvar A., Deboux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussauget O.,
RA Entian K.-D., Fahl H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf U., Jackson D.,
RA Jones L.-M., Kaerst U., Krell T., Kuhn M., Kunst F., Kurapkak G.,
RA Madueno E., Maitounam A., Mata Vicente U., Ng E., Negtari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
CC -1- CATALYTIC ACTIVITY: Shikimate + NADP(+) = 5-dehydroshikimate +
CC NADPH.
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC -1- fourth step.
CC -1- SIMILARITY: Belongs to the shikimate dehydrogenase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL591975; CAC98569.1; -.
CC PIR; A6136; A6136.
CC Listlist; LMO00490; -.
CC HAMAP; MF_00222; -; 1.
CC InterPro; IPR006151; Shikimate_DH.
CC Pfam; PF01488; Shikimate_DH; 1.
CC KMW Aromatic amino acid biosynthesis; Oxidoreductase; NADP;
CC Complete proteome.
SQ SEQUENCE 291 AA; 32161 MW; BDBDF2732CD7D5D CRC64;

```

```

Query Match 40.9%; Score 47; DB 1; Length 291;
Best Local Similarity 72.7%; Pred. No. 3.6;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 2 DIPHTNHKYL 12
Db 74 MPKNTNHHKYL 84

```

```

RESULT 4
PMT1_SCHPO STANDARD; PRT; 330 AA.
ID PMT1_SCHPO

```

AC P40999;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 26-FEB-2003 (Rel. 41, Last annotation update)  
 DE DNA methyltransferase homolog pmt1 (Sp1M.Spot) (M.SpotM1).  
 GN PMT1 OR SPBC19C2.02.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 NX NCBI\_TaxID=4896;  
 RX MEDLINE=972;  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC STRAIN=972;  
 RA MEDLINE=9516638; PubMed=7862522;  
 RA Wilkinson C.R.M., Bartlett R., Nurse P., Bird A.P.;  
 RT "The fission yeast gene pmt1+ encodes a DNA methyltransferase  
 RT homologue."  
 RL Nucleic Acids Res. 23:203-210(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson S., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Voicikert G., Aert R., Robben J., Gymnopreze B.,  
 RA Weljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fitz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Medler H., Wambut R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelarge V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hunt S.M.,  
 RA Lucas M., Roche K., Galliard C., Tallada V.A., Garzon A., Thode G.,  
 RA Duga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cernutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Sipakovski G.V., Ussery D., Barrett B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe."  
 RL Nature 415:871-880(2002).  
 RN [3]  
 RP CHARACTERIZATION.  
 RX MEDLINE=96194447; PubMed=8636983;  
 RA Pinatbasi E., Elliott J., Hornby D.P.;  
 RT "Activation of a yeast pseudo DNA methyltransferase by deletion of a  
 RT single amino acid."  
 RL J. Mol. Biol. 257:804-813(1996).  
 CC -1- FUNCTION: Does not have a cytosine-5 methyltransferase activity  
 CC due to the insertion of a Ser residue between the Pro-Cys motif  
 CC found at the active site of C5 MTases. When this serine is deleted  
 CC it becomes catalytically active and recognizes and methylates the  
 CC sequence CCGATGG.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -1- SIMILARITY: Belongs to the C5-methyltransferase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X82444; CAA57824.1; -

DR EMBL; AL109731; CABS2025.1; -  
 DR PIR; S5390; S5390.  
 DR HSSP; 014717; 1G55.  
 DR REBASE; 2868; M.Spot.  
 DR GeneDB; SPBC19C2.02; -.  
 DR InterPro; IPR001525; C5\_DNA\_meth.  
 DR Pfam; PF00145; DNA\_methylase; 1.  
 DR PRINTS; PR00105; C5METTRFASR.  
 DR PROSITE; PS00094; C5\_MTASE\_1; FALSE\_NEG.  
 DR PROSITE; PS00095; C5\_MTASE\_2; 1.  
 KW DNA-binding; Nuclear protein.  
 FT SITE 81  
 FT SITE 81  
 SQ SEQUENCE 330 AA; 37976 MW; 50A7121FAVCF58A1 CRC64;

Query Match 40.9%; Score 47; DB 1; Length 330;  
 Best Local Similarity 38.9%; Pred. No. 4.2;  
 Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;  
 QY 2 IHPPTNKHVYCEVNG 19  
 DB 108 LPHVNNLPYILIEVNG 125

RESULT 5  
 ID VGLH\_VZVD STANDARD; PRT; 841 AA.  
 AC P09260;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Probable glycoprotein H precursor (Glycoprotein III) (GP111).  
 GN 37.  
 OS Varicella-zoster virus (strain Dumas) (VZV).  
 OS Varicellae; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Varicellovirinae.  
 NX NCBI\_TaxID=10338;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=8630657; PubMed=3018124;  
 RA Davison A.J., Scott J.E.;  
 RT "The complete DNA sequence of varicella-zoster virus."  
 RL J. Gen. Virol. 67:1759-1816(1986).  
 CC -1- SIMILARITY: Belongs to the herpesviruses glycoprotein H family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X04370; CAA27920.1; -

DR PIR; B27341; VGBB37.  
 KW Glycoprotein; Transmembrane; Signal.  
 FT CHAIN 1 841  
 FT CARBOHYD 18 841  
 FT CARBOHYD 18 841  
 FT CARBOHYD 45 45  
 FT CARBOHYD 217 217  
 FT CARBOHYD 317 317  
 FT CARBOHYD 499 499  
 FT CARBOHYD 522 522  
 FT CARBOHYD 760 760  
 FT CARBOHYD 763 763  
 SQ SEQUENCE 841 AA; 93651 MW; 82B247F63CA51948 CRC64;

Query Match 40.9%; Score 47; DB 1; Length 841;  
 Best Local Similarity 50.0%; Pred. No. 12;  
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 QY 2 IHPPTNKHVYCEV 17  
 DB 108 LPHVNNLPYILIEVNG 125

Db 715 LPHDNLKELCYGSGV 730

## RESULT 6

PEL\_ANOGA

ID PEL\_ANOGA STANDARD; PRT; 153 AA.

AC 076217;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Peritrophin-1 precursor.

GN APPEL1

OS Anopheles gambiae (African malaria mosquito).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

CC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.

NCBI\_TaxID=715;

RN [1] SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.

RC STRAIN=G3; TISSUE=Midgut;

RX MEDLINE=9831635; PubMed=9651363;

RA Shen Z., Jacobs-Lorena M.;

RT "A type I peritrophic matrix protein from the malaria vector Anopheles gambiae binds to chitin. Cloning, expression, and characterization.";

RL J. Biol. Chem. 273:17665-17670(1998).

CC -1- FUNCTION: Binds chitin but not cellulose. May be involved in the spatial organization of PM.

CC -1- TISSUE SPECIFICITY: Adult peritrophic membrane.

CC -1- DEVELOPMENTAL STAGE: Expressed in adult but not larval guts, whole pupae or whole bodies minus gut.

CC -1- PTM: Glycosylated.

CC -1- SIMILARITY: Contains 2 chitin-binding type-2 domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AF030431; AAC39127.1; -

DR InterPro; IPR002557; Chitin\_bind\_Pera.

DR Pfam; PF01607; CSM\_14; 2.

DR SMART; SMO0494; CHEBD2; 2.

DR PROSITE; PS050940; CHIT\_BIND\_II\_2.

DR Chitin-binding; Glycoprotein; Repeat; Signal.

KW SIGNAL

FT CHAIN 1 17

FT CHAIN 18 153

FT CHAIN 18 79

FT DOMAIN 92 153

FT DOMAIN 92 63

FT CARBOHYD 63 63

FT SEQUENCE 153 AA; 16819 MW; 2218DA0310476338 CRC64;

SQ

Query Match

Best Local Similarity 40.0%; Score 46; DB 1; Length 153;

Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 2 IPHPTNIHKYIVCE 15

Db 107 IPHPTDCKKYICD 120

Pyrg RICPR

ID PYRG RICPR STANDARD; PRT; 586 AA.

AC 092051;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).

GN PYRG OR RP378.

OS Rickettsia prowazekii.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;

CC

CC

CC

CC Rickettsiaceae; Rickettsiae; Rickettsia.

NCBI\_TaxID=782;

RN [1] SEQUENCE FROM N.A.

RX STRAIN=Madrid B;

RX MEDLINE=9903949; PubMed=9823893;

RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,

RA Sichteritz-Ponten T., Alsmark U.C.M., Podowski R.W., Naeslund A.K.,

RA Eriksson A.-S., Winkler H.E., Kurland C.G.;

RT "The genome sequence of Rickettsia prowazekii and the origin of mitochondria.";

RL Nature 396:133-140(1998).

RN [2] DOMAIN REPI.

RX MEDLINE=20485642; PubMed=11030655;

RA Ogata H., Audic S., Barbe V., Artiguenave F., Fournier P.-E.,

RA Raoul D., Claverie J.-M.;

RT "Selfish DNA in protein-coding genes of Rickettsia.";

RL Science 290:347-350(2000).

CC -1- FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with either L-glutamine or ammonia as the source of nitrogen (By similarity).

CC -1- CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.

CC -1- ENZYME REGULATION: Allosterically activated by GTP, when glutamine is the substrate. Inhibited by CTP (By similarity).

CC -1- SUBUNIT: Homotrimer (By similarity).

CC -1- SIMILARITY: Belongs to the CTP synthase family.

CC -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.

CC -1- SIMILARITY: Contains 1 REPI insert domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AJ235271; GAA14837.1; -

DR PIR; C71695; C71695.

DR HAMAP; MF\_01227; acylpical; 1.

DR InterPro; IPR000991; GATase\_1.

DR InterPro; IPR004468; Pyrg\_synth.

DR InterPro; IPR005728; Rickett\_RPE.

KW Pfam; PF00117; GATase; 1.

DR TIGRFAMs; TIGR00337; Pyrg; 1.

DR TIGRFAMs; TIGR01045; RPE; 1.

DR PROSITE; PS00442; GATase\_TYPE\_I; 1.

KW Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase;

KW Complete proteome.

FT DOMAIN 1 299

FT DOMAIN 300 586

FT DOMAIN 429 473

FT ACT\_SITE 379 379

FT ACT\_SITE 555 555

FT ACT\_SITE 557 557

FT SEQUENCE 586 AA; 66173 MW; FAL4C0879F457A0A CRC64;

SQ

Query Match

Best Local Similarity 40.0%; Score 46; DB 1; Length 586;

Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 2 IPHPTNIHKYIVCE 19

Db 115 IPHPTNIHKYIVCE 132

Pyrg RICPR

ID PYRG RICPR STANDARD; PRT; 891 AA.

AC 092051;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).

GN PYRG OR RP378.

OS Rickettsia prowazekii.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;

CC

CC

01-AUG-1991 (Rel. 19, Created)  
 16-OCT-2001 (Rel. 40, Last sequence update)  
 28-FEB-2003 (Rel. 41, Last annotation update)  
 Protein kinase kind (EC 2.7.1.-).  
 KIN1 OR SPBC46.06.  
 Schizosaccharomyces pombe (Fission yeast).  
 Schizosaccharomycetes; Schizosaccharomycetes;  
 Schizosaccharomycetales; Schizosaccharomycetales;  
 Schizosaccharomycetes.  
 NCBI\_TaxID=4896;  
 [1]  
 SEQUENCE FROM N.A.  
 MEDLINE=91045979; PubMed=2236039;  
 Levin D.E., Bishop J.M.;  
 "A putative protein kinase gene (kin1+) is important for growth  
 polarity in Schizosaccharomyces pombe.";  
 Proc. Natl. Acad. Sci. U.S.A. 87:8272-8276(1990).  
 [2]  
 SEQUENCE FROM N.A.  
 STRAIN=972;  
 MEDLINE=21846401; PubMed=11859360;  
 Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,  
 James K., Jones L., Jones M., Leathers S., McDonald S., McLean J.,  
 Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovich E.,  
 Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 Skelton J., Simmonds M., Squares R., Stevens K.,  
 Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 Woodward J., Volkart G., Aert R., Robben J., Grymptre B.,  
 Melijns I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,  
 Borzym K., Langer I., Beck A., Leinrich H., Reinhardt K., Pohl T.M.,  
 Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,  
 Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaie V., Motier S.,  
 Galbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,  
 Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 Szpakowski G.V., Ussery D., Barrett B.G., Nurse P.;  
 "The genome sequence of Schizosaccharomyces pombe.";  
 Nature 415:871-880(2002).  
 -1- FUNCTION: Probable serine/threonine protein kinase. Important for  
 growth polarity in S.pombe.  
 -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 STRONG, NO YEAST KIN1 AND KIN2.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL: M64999; AAA63577.1; -;  
 EMBL: AL031534; GAA20726.1; -;  
 PIR: A38903; A38903.  
 PIR: T40503; T40503.  
 HSSP: G63450; IAO6.  
 GeneDB SPombe; SPBC46.06; -;  
 InterPro: IPR001772; Kinase\_Cterm.  
 InterPro: IPR000719; Prot\_Kinase.  
 InterPro: IPR008271; Ser\_Thr\_Kin\_AS.  
 InterPro: IPR002290; Ser\_Thr\_Kinase.  
 InterPro: IPR001245; Tyr\_Kinase.  
 Pfam: PF02149; KAI; 1.  
 Pfam: PF00069; kinase; 1.

PRINTS; PRO0109; TYRKINASE.  
 ProDom; PD000001; Prot\_Kinase; 1.  
 SMART; SM00220; S\_TKc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Transferrase; Serine/threonine-protein kinase; ATP-binding.  
 FT DOMAIN 125 395  
 FT NP BIND 131 139  
 FT BINDING 154 154  
 FT ACT\_SITE 266 266  
 FT CONFLICT 141 141  
 FT CONFLICT 247 247  
 FT CONFLICT 620 620  
 FT CONFLICT 777 778  
 FT CONFLICT 891 AA; 96748 MW; 89D9BBD825C0358 CRC64;  
 SQ SEQUENCE  
 Query Match 40.0%; Score 46; DB 1; Length 891;  
 Best Local Similarity 60.0%; Pred. No. 18;  
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 QY 6 TNNHKLVCESVNG 20  
 Db 212 TNSHYWVEFVDG 226  
 ID PNO CRYPV STANDARD; PRT; 1934 AA.  
 AC 0968X;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Pyruvate dehydrogenase [NADP+] (EC 1.2.1.51) (Pyruvate:NADP+  
 oxidoreductase) (CpPNO).  
 GN CPOR.  
 OS Cryptosporidium parvum.  
 OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;  
 OC CRYPTOSPORIDIACEAE; Cryptosporidium.  
 OX NCBI\_TaxID=5807;  
 PN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KSU-1;  
 RX MEDLINE=21219183; PubMed=11319255;  
 RA Rote C., Stejskal P., Zhu G., Keilich J.S., Martin M.;  
 "Pyruvate: NADP oxidoreductase from the mitochondrion of E. coli  
 gracilis and from the apicomplexan Cryptosporidium parvum: A  
 biochemical relic linking pyruvate metabolism in mitochondria and  
 RT biochemodiolate proteins.";  
 RT Mol. Biol. Evol. 18:710-720(2001).  
 CC -1- FUNCTION: May have an important role in respiratory metabolism.  
 Cryptosporidium have a relic mitochondrion with no function in  
 energy metabolism so it is not known if CPOR has a function.  
 CC -1- CATALYTIC ACTIVITY: Pyruvate + CoA + NADP(+) = acetyl-CoA + CO(2)  
 + NADPH.  
 CC -1- COFACTOR: Thiamine pyrophosphate. Flavoprotein that binds one FMN  
 and one FAD per chain.  
 CC -1- SUBUNIT: Homodimer (By similarity).  
 CC -1- DEVELOPMENTAL STAGE: Both sporozoites and intracellular stages of  
 life cycle.  
 CC -1- MISCELLANEOUS: Arose from gene fusion of pyruvate:ferredoxin  
 oxidoreductase and cytochrome-P450 reductase. Gene fusion has only  
 been found in Euglena and Cryptosporidium.  
 CC -1- SIMILARITY: The iron-sulfur centers are similar to those of  
 bacterial-type 4Fe-4S ferredoxins.  
 CC -1- SIMILARITY: Contains 1 flavodoxin-like domain.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sb-sib.ch).

DR EMBL; AF208233; AAK48421.1; -

DR GO; GO:0005739; C:mitochondrion; ISS.

DR GO; GO:0016491; F:oxidoreductase activity; ISS.

DR GO; GO:0045333; P:cellular respiration; ISS.

DR GO; GO:0006090; P:pyruvate metabolism; ISS.

DR InterPro; IPR001450; 4FE4S\_ferredoxin.

DR InterPro; IPR003097; Flav\_binding.

DR InterPro; IPR008254; Flav\_nitox\_synth.

DR InterPro; IPR001094; Flavodoxin-like.

DR InterPro; IPR001226; Flavodoxin.

DR InterPro; IPR001433; Oxid\_FAD/NAD(P).

DR InterPro; IPR002869; POR.

DR InterPro; IPR002880; POR N.

DR InterPro; IPR009014; Transketo\_C-like.

DR pfam; PF00667; FAD\_binding\_1; 1.

DR pfam; PF00037; fer4; 2.

DR pfam; PF00258; Flavodoxin; 1.

DR pfam; PF00175; NAD\_binding\_1; 1.

DR pfam; PF01558; POR; 1.

DR pfam; PF01855; POR N; 1.

DR PRINTS; PR00369; FLAVODOXIN.

DR PRINTS; PR00371; FNCR.

DR PROSITE; PS00198; 4FE4S\_FERREDOXIN; 2.

DR PROSITE; PS00201; FLAVODOXIN; FALSE NEG.

DR PROSITE; PS00902; FLAVODOXIN LIKE; 1.

KW Electron transport; Oxidoreductase; Flavoprotein; NADP; FAD; FMN;

KW Thiamine pyrophosphate; Iron; Iron-sulfur; 4Fe-4S.

FT METAL 719 719

FT METAL 722 722 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).

FT METAL 725 725 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).

FT METAL 729 729 IRON-SULFUR 1 (4FE-4S) (POTENTIAL).

FT METAL 776 776 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).

FT METAL 779 779 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).

FT METAL 782 782 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).

FT METAL 786 786 IRON-SULFUR 2 (4FE-4S) (POTENTIAL).

FT DOMAIN 1288 1438 FAD (ADP PART) (BY SIMILARITY).

FT NP\_BIND 1542 1553 FAD (FLAVIN PART) (BY SIMILARITY).

FT NP\_BIND 1685 1695

FT SEQUENCE 1934 AA; 217556 MW; BCD856F4B2BA3D60 CRC64;

Query Match 40.0%; Score 46; DB 1; Length 1934;

Best Local Similarity 36.8%; Pred. No. 42;

Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 2 IPRPTNHHKYLVCESVNG 20

DB 492 VHHPSYVHKPDVENIKQG 510

RESULT 10

YBS1\_SCHPO STANDARD; PRT; 342 AA.

AC 059681; Q9USX5;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Hypothetical protein C18B5.01 in chromosome II.

GN SPBC18B5.01 OR SPBC29A3.19.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomyces.

OX NCBI\_TaxID=4896;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RX MEDLINE=21848401; PubMed=11859360;

RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,

RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,

RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,

RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,

RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagele K.,

RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,

RA Mooney P., Moutle S., Mungall K., Murphy L., Niblett D., Odell C.,

RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,

RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,

RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,

RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,

RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,

RA Welter J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,

RA Gabel C., Fuchs M., Fritz C., Holzer C., Moestl D., Hilbert H.,

RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,

RA Bieri P., Zimmermann W., Wedler H., Wambut R., Purrelle B.,

RA Coffeau A., Cadieu E., Dreano S., Gloux S., Leleu V., Mottier S.,

RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,

RA Lucas M., Rocher M., Gallardin C., Tallada V.A., Garzon A., Thode G.,

RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,

RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Porzubo S.L.,

RA Cernutti L., Lowe T., Mccombie W.R., Paulsen I., Potashkin J.,

RA Shipakovski G.V., Usery D., Barrell B.G., Nurse P.,

RT The genome sequence of Schizosaccharomyces pombe."

RL Nature 415:871-880(2002).

CC -1- SIMILARITY: Belongs to the cycloisomerase 2 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.sib-sib.ch/announce/>

CC or send an email to license@sb-sib.ch).

CC EMBL; AL022299; CAA18396.1; -

DR EMBL; AL035077; CAA22661.1; -

DR PIR; T40090; T40090.

DR GeneDB; Spombe; SPBC18B5.01; -

KW Hypothetical protein.

FT DOMAIN 163 168 POLY-VAL.

FT SEQUENCE 342 AA; 38220 MW; AE39A39B053F704F CRC64;

Query Match 39.6%; Score 45.5; DB 1; Length 342;

Best Local Similarity 55.6%; Pred. No. 7.6;

Matches 10; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 3 IPRPTNHHKYLVCESVNG 20

DB 130 PHOYTAHKLAVC-SVDLG 146

RESULT 11

CH13\_DROME STANDARD; PRT; 458 AA.

AC Q9W5U2; O17422;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Probable chitinase 3 (BC 3.2.1.14).

GN CH13 OR CG18140.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydriodes; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkelley;

RX MEDLINE=22426071; PubMed=12537574;

RA Hoskins R.A., Smith C.D., Carlson J.W., Carvalho A.B., Halpern A.,

RA Kaminker J.S., Kennedy C., Mungall C.J., Sullivan B.A., Sutton G.G.,

RA Yasuhara T.C., Wakimoto B.T., Myers B.W., Celniker S.E., Rubin G.M.,

RA Karpen G.H.;

RA "Heterochromatic sequences in a Drosophila whole-genome shotgun

```

RT assembly."
RL Genome Biol. 3:RESEARCH0085.1-RESEARCH0085.16(2002).
RL [2]
RP SEQUENCE OF 182-294 FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=9832484; PubMed=9662472;
RA de la Vega H., Specht C.A., Liu Y., Robbins P.W.;
RT "Chitinases are a multi-gene family in Aedes, Anopheles and
  Drosophila."
RL Insect Mol. Biol. 7:233-239(1998).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-
CC acetyl-D-glucosamine polymers of chitin.
CC -1- SIMILARITY: Belongs to chitinase class II (family 18 of glycosyl
CC hydrolases).
CC -1- SIMILARITY: Contains 2 chitin-binding type-2 domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF026502; AAB81860.1; -.
DR FlyBase; FBgn002701; Chit3.
DR InterPro; IPR002557; Chitin bind_Pera.
DR InterPro; IPR001223; Glyco_hydro_18.
DR InterPro; IPR001579; Glyco_hydro_18AS.
DR Pfam; PF01607; CBM_14; 2.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR ProDom; PD000471; Glyco_hydro_18; 1.
DR SMART; SM00494; ChitBD2; 2.
DR SMART; SM00636; Glyco_18; 1.
DR PROSITE; PS00940; CHIT BIND II; 2.
DR PROSITE; PS01095; CHITINASE_18; PULSE NEG.
KM Hydrolyase; Glycosidase; Chitin degradation; Chitin-binding;
KM Multigene family; Repeat.
FT DOMAIN 5 58 CHITIN-BINDING TYPE-2 1.
FT ACT SITE 74 128 CHITIN-BINDING TYPE-2 2.
FT ACT SITE 295 295 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 458 AA; 52330 MW; 4A063190B7E96248 CRC64;

Query Match 39.1%; Score 45; DB 1; Length 458;
Best Local Similarity 53.8%; Pred. No. 13;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 IPHPTNHYKYLVC 14
DB 83 VPPGNCCKYLFC 95

RESULT 12
CHIT_BRUMA STANDARD; PRT; 504 AA.
ID CHIT_BRUMA
AC P29030;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Endochitinase precursor (EC 3.2.1.14) (MFI antigen).
OS Brugia malayi (Filarial nematode worm).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Brugia.
OX NCBI_TaxID=6279;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=92179220; PubMed=1542646;
RA Fukumori J.A., Lane W.S., Smith R.P., Plessens W.F., Perler F.B.;
RT "Transmission-blocking antibodies recognize microfilarial chitinase
  in brugian lymphatic filariasis."
RL Proc. Natl. Acad. Sci. U.S.A. 89:1548-1552(1992).
CC -1- FUNCTION: Microfilarial chitinase, which may function to degrade
  chitin-containing structures in the micro-filaria or in its

```

```

CC mosquito vector during parasite development and transmission.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-
CC acetyl-D-glucosamine polymers of chitin.
CC -1- DEVELOPMENTAL STAGE: The appearance of the MFI antigen correspond
CC with the onset of the parasite's ability to infect the mosquito.
CC -1- PTM: O-glycosylated.
CC -1- MISCELLANEOUS: Known to bind calcium.
CC -1- SIMILARITY: Belongs to chitinase class II (family 18 of glycosyl
CC hydrolases).
CC -1- SIMILARITY: Contains 1 chitin-binding type-2 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M73689; AAA27854.1; -.
DR PIR; A38221; A38221.
DR InterPro; IPR002557; Chitin bind_Pera.
DR InterPro; IPR001223; Glyco_hydro_18.
DR InterPro; IPR001579; Glyco_hydro_18AS.
DR Pfam; PF01607; CBM_14; 1.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR ProDom; PD000471; Glyco_hydro_18; 1.
DR SMART; SM00494; ChitBD2; 1.
DR SMART; SM00636; Glyco_18; 1.
DR PROSITE; PS00940; CHIT BIND II; 1.
DR PROSITE; PS01095; CHITINASE_18; 1.
KM Hydrolyase; Glycosidase; Chitin degradation; Chitin-binding; Signal;
KM Antigen; Repeat; Glycoprotein; Calcium-binding.
FT SIGNAL 1 22
FT CHAIN 23 504 ENDOCHITINASE.
FT DOMAIN 23 400 CATALYTIC.
FT DOMAIN 401 448 SER/THR-RICH (LINKER).
FT DOMAIN 407 448 3 X 14 AA APPROXIMATE TANDEM REPEATS.
FT DOMAIN 448 504 CHITIN-BINDING TYPE-2.
FT ACT SITE 148 148 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 504 AA; 55971 MW; A78BE7BF8E83709B CRC64;

Query Match 39.1%; Score 45; DB 1; Length 504;
Best Local Similarity 50.0%; Pred. No. 14;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 PEPPTNHYKYLVC 14
DB 459 PAPTCHLPIQC 470

RESULT 13
RS10_FUSNN STANDARD; PRT; 103 AA.
ID RS10_FUSNN
AC Q8RIF4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S10.
GN RPSJ OR FNI646.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteriia; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Battacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haseikorn R.,
RA Fonstein M., Kyripides N., Overbeek R.;

```

```

RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586."
RT J. Bacteriol. 184:2005-2018(2002).
CC -1- FUNCTION: Involved in the binding of tRNA to the ribosomes (By
CC similarity).
CC -1- SIMILARITY: Belongs to the S10P family of ribosomal proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE010471; AAL93761.1; -.
DR HAMAP; MF_00508; -.
DR InterPro; IPR001848; Ribosomal_S10_b.
DR InterPro; IPR005731; Ribosomal_S10_b.
DR Pfam; PF00338; Ribosomal_S10; 1.
DR PRINTS; PR00971; RIBOSOMAL_S10.
DR PRODOM; PD001272; Ribosomal_S10; 1.
DR TIGRFAIMS; TIGR01049; rps1_bact; 1.
DR PROSITE; PS00361; RIBOSOMAL_S10; 1.
KM Ribosomal protein, complete proteome.
SQ SEQUENCE 103 AA; 11541 MW; EFSNA895B8F63DCA CRC64;

QY Query Match 38.3%; Score 44; DB 1; Length 103;
Best Local Similarity 52.9%; Pred. No. 3.5;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Db 2 IHPPTNIHKYLVCSVN 18
40 MFLPTKIRKXTVLRVSH 56

RESULT 14
EIA_ADEC2 STANDARD; PRT; 171 AA.
AC P35981;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Early EIA 20 kDa protein.
OS Canine adenovirus type 2.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=10514;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=90163565; PubMed=2623943;
RA Spilbey N., McGilroy R.S., Cavanagh H.M.A.;
RT "Identification and nucleotide sequence of the early region 1 from
RT canine adenovirus types 1 and 2."
RL Virus Res. 14:241-256(1989).
DR PIR; B60010; B60010.
KM Transcription regulation: Early protein.
SQ SEQUENCE 171 AA; 18942 MW; 2527EC1338062FB0 CRC64;

QY Query Match 38.3%; Score 44; DB 1; Length 171;
Best Local Similarity 30.0%; Pred. No. 6.2;
Matches 9; Conservative 4; Mismatches 5; Indels 12; Gaps 1;

Db 2 IHPPTNIHKYLV-----CESVNG 19
6 VPAPRNLDHYVLELSEWHPDCLDCEYPNG 35

RESULT 15
VSP1_AGKCO STANDARD; PRT; 231 AA.
AC P09872;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)

```

```

DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Anceid (EC 3.4.21.74) (Venombin A) (Protein C activator) (ACC-C).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodonta; Squamata; Sclerozoa; Serpentes; Colubroidae;
OC Viperidae; Crotalinae; Agkistrodon.
OX NCBI_TaxID=8713;
RN [1]
RP SEQUENCE.
RC TISSUE=Venom;
RX MEDLINE=89229065; PubMed=2653426;
RA McMullen B.A., Fujikawa K., Kiesel W.;
RT "Primary structure of a protein C activator from Agkistrodon
RT contortrix contortrix venom."
RL Biochemistry 28:674-679(1989).
RN [2]
RP SEQUENCE OF 1-63.
RC TISSUE=Venom;
RX MEDLINE=87308291; PubMed=3624272;
RA Kiesel W., Kondo S., Smith K.J., McMullen B.A., Smith L.F.;
RT "Characterization of a protein C activator from Agkistrodon
RT contortrix contortrix venom."
RL J. Biol. Chem. 262:12607-12613(1987).
CC -1- FUNCTION: Thrombin-like snake venom serine protease. Cleaves
CC fibrinogen to form fibrin, and release fibrinopeptide A. The
CC incapable of being cross-linked, forming easily dispersible
CC clots. Activates protein C.
CC -1- CATALYTIC ACTIVITY: Selective cleavage of Arg-Xaa bond in
CC fibrinogen, to form fibrin, and release fibrinopeptide A. The
CC specificity of further degradation of fibrinogen varies with
CC species origin of the enzyme.
CC -1- SUBCELLULAR LOCATION: Secreted (Potential).
CC -1- TISSUE SPECIFICITY: Expressed by the venom gland.
CC -1- SIMILARITY: Belongs to peptidase family S1. Snake venom subfamily.
DR PIR; A60468; A60468.
DR HSP; P00763; IDPO.
DR MEROPS; S01.178.00.
DR InterPro; IPR009003; Cys_ser_trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR00314; Peptidase_S1A.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN_DOM; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; FALSE_NEG.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KM Hydroxylase, Serine protease, glycoprotein.
FT ACT_SITE 40 40
FT ACT_SITE 85 85
FT ACT_SITE 177 177
FT DISULFID 7 138
FT DISULFID 25 41
FT DISULFID 73 229
FT DISULFID 117 183
FT DISULFID 149 168
FT DISULFID 173 192
FT CARBOHYD 21 21
FT CARBOHYD 76 76
FT CARBOHYD 129 129
SQ SEQUENCE 231 AA; 25106 MW; DC3D7CB8601EC52B CRC64;

QY Query Match 38.3%; Score 44; DB 1; Length 231;
Best Local Similarity 40.9%; Pred. No. 8.6;
Matches 9; Conservative 4; Mismatches 7; Indels 2; Gaps 1;

Db 1 DIPPTNIH--KYLVCESVNG 20
134 DVPHCANINIDYAVQAAAYKG 155

Search completed: March 22, 2004, 06:52:55
Job time : 2.76709 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:39:53 ; Search time 4.02446 Seconds

(without alignments)  
1568.003 Million cell updates/sec

Title: US-09-662-293-3

Perfect score: 115  
Sequence: 1 DIPHTHICKYVCEVNGG 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	109	94.8	555	5	Q9U6R7
2	57	49.6	523	5	Q8MNX4
3	56	48.7	1635	5	Q8MNX4
4	55	47.8	517	10	Q9S792
5	55	47.8	437	10	Q9XR74
6	54	47.0	563	5	Q8T015
7	54	47.0	563	5	Q9VST9
8	54	47.0	2382	5	Q8B119
9	54	47.0	2786	5	Q9VSU2
10	53.5	46.5	66	9	Q03967
11	53	46.1	260	13	Q8UXY1
12	53	46.1	525	5	Q84079
13	52	45.2	727	5	Q21139
14	51	44.3	721	10	Q94DE8
15	51	44.3	721	10	Q94DE8
16	49	42.6	237	5	Q9VR79

17	49	42.6	504	5	Q9NGX8	Q9ngk8 wuchereria
18	49	42.6	557	10	Q7XUD4	Q7xud4 oryza sativ
19	49	42.6	647	15	Q9Q839	Q9q839 human t-lym
20	49	42.6	651	15	Q56227	Q56227 human t-lym
21	49	42.6	796	5	Q9VTR4	Q9vtr4 drosophila
22	49	42.6	796	5	Q8MRG9	Q8mr9 drosophila
23	49	42.6	1273	15	Q9Q8A1	Q9q8a1 human t-lym
24	49	42.6	1462	15	Q56228	Q56228 human t-lym
25	48	41.7	242	10	Q8W126	Q8w126 solanum tub
26	47.5	41.3	616	5	Q8MQ14	Q8mq14 drosophila
27	47.5	41.3	1039	5	Q9VP14	Q9vp14 drosophila
28	47.5	41.3	2833	5	Q9VP13	Q9vp13 drosophila
29	47	40.9	258	13	Q8AY80	Q8ay80 trimeresuru
30	47	40.9	258	13	Q8AY78	Q8ay78 trimeresuru
31	47	40.9	290	10	Q9ARD1	Q9ard1 lycopersico
32	47	40.9	486	5	Q8TSC4	Q8tsc4 aedes aegypt
33	47	40.9	841	12	Q9DWY1	Q9dwy1 human herpe
34	47	40.9	841	12	Q9DX81	Q9dx81 human herpe
35	47	40.9	841	12	Q98VL1	Q98vl1 human herpe
36	47	40.9	841	12	Q80AC3	Q80ac3 human herpe
37	47	40.9	841	12	Q80AC1	Q80ac1 human herpe
38	47	40.9	852	12	Q66030	Q66030 cercopithec
39	47	40.9	894	5	Q81511	Q81511 plasmodium
40	47	40.9	1290	5	Q9VTR8	Q9vtr8 drosophila
41	47	40.9	1324	5	Q8SYK2	Q8syk2 drosophila
42	46.5	40.4	229	11	Q8C466	Q8c466 mus musculu
43	46	40.0	183	10	Q7XZC8	Q7xzc8 nicotiana t
44	46	40.0	452	3	Q8UOK6	Q8uok6 melancarpu
45	46	40.0	503	5	Q9NAR8	Q9nar8 branchiosto

## ALIGNMENTS

RESULT 1  
ID Q9U6R7 PRELIMINARY; PRT; 555 AA.  
AC Q9U6R7;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE 98kDa HDM allergen.  
OS Dermatophagoides farinae (House-dust mite).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
OC Acariformes; Sarcoptiformes; Astigmata; Psoroptidia; Analgoidea;  
OC Pyroglyphidae; Dermatophagoides.  
OX NCBI\_TaxID=6954;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Weber E.R., Hunter S., Steedman K., McCall C.;  
RT "Cloning and Characterization of a 98 kDa Allergen from  
RT Dermatophagoides farinae";  
RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF178772; AAD52672.1; -  
DR GO; GO:000576; C:extracellular; IEA.  
DR GO; GO:0008061; P:chitin binding; IEA.  
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR GO; GO:0006030; P:chitin metabolism; IEA.  
DR InterPro; IPR002557; Chitin bind PERA.  
DR InterPro; IPR001223; Glyco\_hydro\_18.  
DR InterPro; IPR001579; Glyco\_hydro\_18AS.  
DR Pfam; PF00704; Glyco\_hydro\_18; 1.  
DR ProDom; PD000471; Glyco\_hydro\_18; 1.  
DR SMART; SM00494; ChnBD; 1.  
DR SMART; SM00636; Glyco\_18; 1.  
DR PROSITE; PS01095; CHITINASE\_18; 1.  
KW Glycosidase; Hydrolase.  
SQ SSQDNCE 555 AA; 65238 MM; 0E45641AA59B30B CRC64;  
Query Match 94.8%; Score 109; DB 5; Length 555;  
Best Local Similarity 95.0%; Pred. No. 1.4e-09;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;



QY 1 DIPHTNINHYLVCSVNG 20  
DB 512 DIPHTNINHYLVCSVNG 531

RESULT 2  
Q8MNX4 PRELIMINARY; PRT; 523 AA.

AC Q8MNX4;  
DT 01-OCT-2002 (TREMBlrel. 22, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
DE Hypothetical protein.  
GN T11F1.7.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.; Marmley P.;  
RT "The sequence of C. elegans cosmid T11F1."  
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF098996; AM34815.1; -  
DR WormRep; T11F1.7; CE23978.  
DR GO; GO:0016020; Cmembrane; IEA.  
DR GO; GO:005006; F:epidermal growth factor receptor activity; IEA.  
DR InterPro; IPR000494; EGFR\_L domain.  
DR Pfam; PF01030; Recep\_L domain; 2.  
KM Hypothetical protein.  
SQ SEQUENCE 523 AA; 60198 MW; CODCF493F0C34ED5 CRC64;

Query Match 49.6%; Score 57; DB 5; Length 523;  
Best Local Similarity 55.0%; Pred. No. 0.65;  
Matches 11; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 DIPHTNINHYLVCSVNG 20  
DB 193 DIPHTNINHYLVCSVNG 212

RESULT 3  
Q17412 PRELIMINARY; PRT; 1635 AA.

AC Q17412;  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JUN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Probable chitinase 2 (EC 3.2.1.14).  
GN CHT2.  
OS Aedes aegypti (Yellowfever mosquito).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Aedes.  
OX NCBI\_TaxID=7159;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=98324849; PubMed=9662472;  
RA de la Vega H., Specht C.A., Liu Y., Robbins P.W.;  
RT "Chitinases are a multi-gene family in Aedes, Anopheles and Drosophila."  
RL Insect Mol. Biol. 7:233-239 (1998).  
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES OF N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.

CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL HYDROLASES).  
CC EMBL; AF026492; AAB81850.1; -  
DR PIR; T14075; T14075.  
DR GO; GO:000576; C:extracellular; IEA.  
DR GO; GO:0008061; F:chitin binding; IEA.  
DR GO; GO:0008943; F:endochitinase activity; IEA.  
DR GO; GO:0016798; F:hydrolyase activity, acting on glycosyl bonds; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR GO; GO:0006032; P:chitin catabolism; IEA.  
DR InterPro; IPR002557; Chitin\_bind\_Pera.  
DR InterPro; IPR001223; Glyco\_hydro\_18.  
DR InterPro; IPR001579; Glyco\_hydro\_18AS.  
DR Pfam; PF01607; CBM\_14; 3.  
DR Pfam; PF00704; Glyco\_hydro\_18; 3.  
DR ProDom; PD000471; Glyco\_hydro\_18; 3.  
DR SMART; SM00494; ChnBD2; 3.  
DR SMART; SM00636; Glyco\_18; 3.  
DR PROSITE; PS01095; CHITINASE\_18; 3.  
KM Hydrolyase; Glycosidase; Chitin degradation; Glycoprotein;  
KM Multigene family.  
FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 463 463 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 749 749 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 890 890 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1338 1338 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1479 1479 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1635 AA; 185993 MW; EA116F83AAC129FA CRC64;

Query Match 48.7%; Score 56; DB 5; Length 1635;  
Best Local Similarity 50.0%; Pred. No. 3.2;  
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 IPHPTNINHYLVCS 15  
DB 1134 VPHPTDNKHYTCO 1147

RESULT 4  
Q9S792 PRELIMINARY; PRT; 517 AA.

AC Q9S792;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Calcium/calmodulin dependent protein kinase.  
GN CCAK OR CCAK-1.  
OS Nicotiana tabacum (Common tobacco).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
OC Lamiales; Solanales; Solanaceae; Nicotiana.  
OX NCBI\_TaxID=4097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. XANTHI; TISSUE=Anther;  
RA Liu Z., Poovaiah B.W.;  
RT "Regulated expression of an anther-specific calcium/calmodulin dependent protein kinase causes male sterility in plant."  
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. XANTHI; TISSUE=anther;  
RA Liu Z.H., Xia M., Poovaiah B.W.;  
RT "Chimeric calcium/calmodulin-dependent protein kinase in tobacco: differential regulation by calmodulin isoforms."  
RL Plant Mol. Biol. 0:0-0 (1998).  
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
DR EMBL; U38446; AAF21450.1; -  
DR EMBL; AF087813; AAD52092.1; -  
DR HSPB; Q63450; 1A06.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0005509; F:calcium ion binding; IEA.

DR GO:0004674; F:protein serine/threonine kinase activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006468; P:protein amino acid phosphorylation; IEA.  
 DR InterPro: IPR002048; EF-hand.  
 DR InterPro: IPR000719; Prot\_kinase.  
 DR InterPro: IPR001125; Recoverin.  
 DR InterPro: IPR002290; Ser\_thr\_kinase.  
 DR InterPro: IPR008271; Ser\_thr\_pkin\_AS.  
 DR Pfam: PF00036; ehand; 3.  
 DR Pfam: PF00069; pkinase; 1.  
 DR PRINTS: PR00450; RECOVERIN.  
 DR ProDom: PD000012; EF-hand; 1.  
 DR ProDom: PD000001; Prot\_kinase; 1.  
 DR SMART: SM00054; EFh; 3.  
 DR SMART: SM00220; S\_TKc; 1.  
 DR SMART: SM00018; EF\_HAND; 3.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
 SQ SEQUENCE 517 AA; 57624 MW; BB6706E29AFB349D CRC64;

Query Match 47.8%; Score 55; DB 10; Length 517;  
 Best Local Similarity 58.8%; Pred. No. 1.4;  
 Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 3 PHPTNKHLYVCEVNG 19  
 DB 97 PHPNVHLYVCEDPG 113

RESULT 5  
 Q9XEU4 PRELIMINARY; PRT; 517 AA.  
 AC Q9XEU4;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Calcium/calmodulin-dependent protein kinase.  
 CC CAMK OR CCAMK-2.  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
 OC Lamiales; Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxID=4097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. XANTHI; TISSUE=Anther.  
 RA Wang W., Liu Z.H., Xia M., Poovaiah B.W.;  
 RT "Chimeric calcium/calmodulin-dependent protein kinase in tobacco:  
 RT differential regulation by calmodulin isoforms."  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. XANTHI; TISSUE=Anther;  
 RA Liu Z.H., Xia M., Poovaiah B.W.;  
 RT "Chimeric calcium/calmodulin-dependent protein kinase in tobacco:  
 RT differential regulation by calmodulin isoforms."  
 RL Plant Mol. Biol. 0:0-0(1998).  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL: AF145593; AAD28791.1; -. JOINED.  
 DR EMBL: AF145592; AAD28791.1; -. JOINED.  
 DR HSSP: O63450; 1A06.  
 DR GO:0005524; F:ATP binding; IEA.  
 DR GO:0005509; F:calcium ion binding; IEA.  
 DR GO:0004674; F:protein serine/threonine kinase activity; IEA.  
 DR GO:0016740; F:transferase activity; IEA.  
 DR GO:0006468; P:protein amino acid phosphorylation; IEA.  
 DR InterPro: IPR002048; EF-hand.  
 DR InterPro: IPR000719; Prot\_kinase.  
 DR InterPro: IPR001125; Recoverin.  
 DR InterPro: IPR002290; Ser\_thr\_kinase.

DR InterPro: IPR008271; Ser\_thr\_pkin\_AS.  
 DR Pfam: PF00036; ehand; 3.  
 DR Pfam: PF00069; pkinase; 1.  
 DR PRINTS: PR00450; RECOVERIN.  
 DR ProDom: PD000012; EF-hand; 1.  
 DR ProDom: PD000001; Prot\_kinase; 1.  
 DR SMART: SM00054; EFh; 3.  
 DR SMART: SM00220; S\_TKc; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
 SQ SEQUENCE 517 AA; 57881 MW; 48F8361E2E80A61 CRC64;

Query Match 47.8%; Score 55; DB 10; Length 517;  
 Best Local Similarity 58.8%; Pred. No. 1.4;  
 Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 3 PHPTNKHLYVCEVNG 19  
 DB 97 PHPNVHLYVCEDPG 113

RESULT 6  
 Q8TO15 PRELIMINARY; PRT; 437 AA.  
 AC Q8TO15;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE GH28017P.  
 GN TBQUTLA OR CG4821 OR CG4948 OR CG18403.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkely;  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champagne M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
 RA Gonzalez M., Guatlin H., Li P., Liao G., Miranda A., Mungall C.T.,  
 RA Munoz J., Pacleb J., Paragas V., Park S., Pounenavong S., Wan K.,  
 RA Yu C., Lewis S.B., Rubin G.M., Ceiniker S.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY069235; AAL39380.1; -.  
 DR FlyBase: FBgn0023479; Tegulla.  
 DR GO:0005576; C:extracellular; IEA.  
 DR GO:0008061; P:chitin binding; IEA.  
 DR GO:0006030; P:chitin metabolism; IEA.  
 DR InterPro: IPR02557; Chitin\_bind\_Pera.  
 DR Pfam: PF01607; CBM\_14; 5.  
 DR SMART: SM00494; ChIBD2; 5.  
 SQ SEQUENCE 437 AA; 49481 MW; 1D2D73B1E7E5CF20 CRC64;

Query Match 47.0%; Score 54; DB 5; Length 437;  
 Best Local Similarity 66.7%; Pred. No. 1.7;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 PHPTNKHLYVCEVNG 14  
 DB 98 PHPNVHLYVCEDPG 109

RESULT 7  
 Q9VST9 PRELIMINARY; PRT; 563 AA.  
 AC Q9VST9;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE CG4821 protein.

GN TEQUILA OR CG4821 OR CG4948 OR CG18403.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NC NCB1\_TaxID=7227;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amaratunga P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Morten J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abiri J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bertan B.P., Bhattacharya D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck U., Brokstein P., Brotter P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
 RA Glöckle A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibbegan C.,  
 RA Jafali M., Kalish F., Karpen G.H., Ke Z., Kennison J.D., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milhina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclik J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao S., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amaratunga P.G., Brandon R.C., Rogers Y.,  
 RA Barton J., An H., Baldwin D., Barton J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferreira S., Frise E., Galie R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibbegan C., Jafali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuno J.,  
 RA Paclik J., Paragass V., Park S., Patel S., Pfeiffer B.,  
 RA Phoonanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Starckel M., Strong R., Svitek R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RP SEQUENCE FROM N.A.  
 RA Mista S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminer J.S., Prochuk S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
 RA Clamp W., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Milburn J., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,

RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 [5]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF003553; AAF50322.2; -  
 DR FlyBase; FBgn0023479; Tequila.  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0008061; F:chitin binding; IEA.  
 DR GO; GO:0008061; F:chitin binding; IEA.  
 DR GO; GO:0006030; P:chitin metabolism; IEA.  
 DR InterPro; IPR002557; Chitin\_bind\_perA.  
 DR Pfam; PF01607; CBM\_14; 7.  
 DR SMART; SM00494; ChIBD2; 7.  
 SQ SEQUENCE 563 AA; 61963 MW; 9AC106E0F4913F93 CRC64;  
 Query Match 47.0%; Score 54; DB 5; Length 563;  
 Best Local Similarity 66.7%; Pred. No. 2.2;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 3 PHEPNIHKLIVC 14  
 Db 224 PHEPDVHKYLRC 235  
 ID 09B119 PRELIMINARY; PRT; 2382 AA.  
 AC 09B119;  
 DT 01-JUN-2001 (TEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
 DE GRAAL2 protein precursor.  
 GN TEQUILA OR GRAAL OR CG4821 OR CG4948 OR CG18403.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NC NCB1\_TaxID=7227;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RA Manier A.I., Medzhitov R., Janeway C.A., Lanot R., Zachary D.,  
 RA Capovilla M., Lagueux M.;  
 RT "Graal a Drosophila gene coding for several mosaic serine proteases.";  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 DR EMBL; AJ309005; CAC35209.1; -  
 DR HSP; P00750; IRTF  
 DR FlyBase; FBgn0023479; Tequila.  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0008061; F:chitin binding; IEA.  
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0005044; F:scavenger receptor activity; IEA.  
 DR GO; GO:0004295; F:trypsin activity; IEA.  
 DR GO; GO:0006030; P:chitin metabolism; IEA.  
 DR GO; GO:0006030; P:chitin metabolism; IEA.  
 DR InterPro; IPR002557; Chitin\_bind\_perA.  
 DR InterPro; IPR009003; Cys\_ser\_trypsin.  
 DR InterPro; IPR002172; ID1\_receptor\_A.  
 DR InterPro; IPR001254; Peptidase\_S1.  
 DR InterPro; IPR001314; Peptidase\_S1A.  
 DR InterPro; IPR001190; Srcr\_receptor.  
 DR Pfam; PF01607; CBM\_14; 15.  
 DR Pfam; PF00057; Id1\_recept\_a; 2.  
 DR Pfam; PF00530; SRCR; 2.  
 DR Pfam; PF00089; trypsin; 1.



DR PRINTS; PR00261; LDLRECEPTOR.  
 DR SMART; SM00494; ChED2; 15.  
 DR SMART; SM00192; LDLA; 2.  
 DR SMART; SM00202; SR; 2.  
 DR SMART; SM00020; TRYP\_SPE; 1.  
 DR PROSITE; PS00152; ATPASE ALPHA\_BETA; 1.  
 DR PROSITE; PS01209; LDLRA\_1; 1.  
 DR PROSITE; PS50068; LDLRA\_2; 2.  
 DR PROSITE; PS00420; SRCR\_1; 2.  
 DR PROSITE; PS50287; SRCR\_2; 2.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 DR Hydrolase; Protease; Serine protease.  
 SQ SEQUENCE 2786 AA; 308362 MW; 2EED7A7DA502C76 CRC64;

Query Match 47.0%; Score 54; DB 5; Length 2786;  
 Best Local Similarity 66.7%; Pred. No. 12;  
 Matches 8; Conservative 2; Mismatches 0; Gaps 0;

Qy 3 PPHPTNHYKLYVC 14  
 |||:|||||  
 Db 224 PPHPTNHYKLYVC 235

RESULT 10  
 003967 PRELIMINARY; PRT; 66 AA.

AC 003967; (TREMBlrel. 04, Created)  
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE EXICIONASE.  
 GN XIS OR GP66.  
 OS Bacteriophage phigle.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.  
 OX NCBI\_TaxID=52979;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97225795; PubMed=9073065;  
 RA Kodaira K.I., Oki M., Kakikawa M., Watanabe N., Hirakawa M.,  
 RA Yamada K., Takeo A.;  
 DR EMBL; X90510; CAA62091.1; -.  
 SQ SEQUENCE 66 AA; 7557 MW; 4FDA25D9440CD1E4 CRC64;

Query Match 46.5%; Score 53.5; DB 9; Length 66;  
 Best Local Similarity 50.0%; Pred. No. 0.27;  
 Matches 9; Conservative 4; Mismatches 2; Indels 3; Gaps 1;

Qy 1 DIPPTNI--HKYLVCE 15  
 |||:|||||  
 Db 37 DVFPPTNIASYHFLACD 54

RESULT 11  
 08UVX1 PRELIMINARY; PRT; 260 AA.

AC 08UVX1; (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Thrombin-like serine protease (Thrombin-like enzyme)  
 DE (EC 3.4.21.7).  
 GN TLE.

OS Agkistrodon caliginosus (Korean viper) (Gloydius ussuriensis).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae;  
 OC Viperidae; Crotalinae; Gloydius.  
 OX NCBI\_TaxID=356711;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=venom gland;  
 RA Zhao Y., Fang K., Sun K.;  
 RT "cDNA for thrombin-like serine protease from venom gland of  
 RT Agkistrodon ussuriensis.";  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Yang Q., An L., Su Z.;  
 RT "cDNA sequence analysis and expression in yeast Pichia pastoris of a  
 RT thrombin-like enzyme.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 DR EMBL; AF336126; AAL68708.1; -.  
 DR EMBL; AF370124; AAM46086.1; -.

DR HSP; P00761; IAN1.  
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0004283; F:plasmin activity; IEA.  
 DR GO; GO:0004295; F:trypsin activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR009003; Cys Ser trypsin.  
 DR InterPro; IPR001254; Peptidase\_S1.  
 DR InterPro; IPR001314; Peptidase\_S1A.  
 DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR PROSITE; PS50240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolase; Protease; Serine protease.  
 SQ SEQUENCE 260 AA; 28530 MW; 00514BE487B6F84C CRC64;

Query Match 46.1%; Score 53; DB 13; Length 260;  
 Best Local Similarity 45.5%; Pred. No. 1.4;  
 Matches 10; Conservative 5; Mismatches 5; Indels 2; Gaps 1;

Qy 1 DIPPTNIH--KYLCESYNGC 20  
 |||:|||||  
 Db 161 DVPHCANINILDYVCGAAGC 182

RESULT 12  
 044079 PRELIMINARY; PRT; 525 AA.

AC 044079; (TREMBlrel. 06, Created)  
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Chitinase.

DE Chitinase.  
 GN AGCHT-1.  
 OS Anopheles gambiae (African malaria mosquito).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Nematocera; Anopheles.  
 OX NCBI\_TaxID=7165;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Gut;  
 RX MEDLINE=98030563; PubMed=9360958;  
 RA Shen Z., Jacobs-Lorena M.;  
 RT "Characterization of a novel gut-specific chitinase gene from the  
 RT human malaria vector Anopheles gambiae.";  
 RL U. Biol. Chem. 272:28895-28900(1997).

DR EMBL; AF008575; AAB87764.1; -.  
 DR PIR; T44445; T44445.  
 DR GO; GO:000576; C:extracellular; IEA.  
 DR GO; GO:0008061; F:chitin binding; IEA.  
 DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.

DR GO:0005975; P:carbohydrate metabolism; IEA.  
 DR GO:0006030; P:chitin metabolism; IEA.  
 DR InterPro: IPR002557; Chitin bind. per.  
 DR InterPro: IPR001223; Glyco\_hydro\_18.  
 DR InterPro: IPR001579; Glyco\_hydro\_18AS.  
 DR Pfam: PF01607; GEM\_14; 1.  
 DR Pfam: PF00704; Glyco\_hydro\_18; 1.  
 DR ProDom: PD000471; Glyco\_hydro\_18; 1.  
 DR SMART: SM00494; ChEBD2; 1.  
 DR SMART: SM00636; Glyco\_18; 1.  
 DR PROSITE: PS01095; CHITINASE\_18; 1.  
 KM Glycoisidase; Hydrolase.  
 SQ SEQUENCE 525 AA; 5721 MW; 3234360EFFF36165 CRC64;

Query Match 46.1%; Score 53; DB 5; Length 525;  
 Best Local Similarity 53.8%; Pred. No. 3;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 IPRPTNIXYLVC 14  
 Db 478 VPHPTNCRYYIC 490

RESULT 13  
 Q21139 PRELIMINARY; PRT; 727 AA.

AC 021139;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Synaplogensis abnormal protein 1.  
 GN K02E10.8 OR SYG-1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_Taxid=6239;  
 RN 1;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA Wilson R.;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:12012-2018(1998).  
 RN 12;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Wu X., Leimbach D.;  
 RT "The sequence of C. elegans cosmid K02E10.";  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 RN 13;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U40942; AAC47074.5; -.  
 DR PIR: T16525; T16525.  
 DR Wormpep: K02E10.8; CB33930.  
 DR InterPro: IPR003599; Ig-like.  
 DR InterPro: IPR007110; Ig-like.  
 DR InterPro: IPR003598; Ig\_c2.  
 DR Pfam: PF00047; Ig; 4.  
 DR SMART: SM00409; Ig; 4.  
 DR SMART: SM00408; IGG2; 4.  
 DR PROSITE: PSS0835; IG\_LIKE; 5.  
 SQ SEQUENCE 727 AA; 80863 MW; C623B14C92205E2A CRC64;

Query Match 45.2%; Score 52; DB 5; Length 727;  
 Best Local Similarity 60.0%; Pred. No. 6.3;  
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 IPRPTNIXYLVCES 16  
 |||:||||:|

Db 234 IPRPDDHKYLICIS 248

RESULT 14  
 Q77229 PRELIMINARY; PRT; 260 AA.  
 ID Q77229;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Serine protease.  
 OS Bothrops jararacussu (Jararacussu).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;  
 OC Viperidae; Crotalinae; Bothrops.  
 OX NCBI\_Taxid=8726;  
 RN 1;  
 RP SEQUENCE FROM N.A.  
 RA Kashima S., Roberto P.G., Soares A.M., Astolfi-Filho S., Pereira J.O.,  
 RA Giglio J.R., Franca S.C.;  
 RT "Analysis of Bothrops jararacussu venomous gland transcriptome with  
 RT structural and functional categories: gene expression profile of  
 RT highly expressed phospholipases A2.";  
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY251282; AAP42416.1; -.  
 KW Protease.  
 SQ SEQUENCE 260 AA; 28654 MW; 9E633F098E51F8CF CRC64;

Query Match 44.3%; Score 51; DB 13; Length 260;  
 Best Local Similarity 45.5%; Pred. No. 3.1;  
 Matches 10; Conservative 4; Mismatches 6; Indels 2; Gaps 1;

QY 1 DIPPTNIR--XYLVCSYNGG 20  
 Db 161 DVPHCMINLDYVCRAHGG 182

RESULT 15

Q94DB8 PRELIMINARY; PRT; 721 AA.  
 ID Q94DB8;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE P0683F02.4 protein (OJ1402.H07.12 protein).  
 GN P0683F02.4 OR OJ1402.H07.12.  
 OS Oryza sativa (Rice), and  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_Taxid=4530, 39947;  
 RN 11;  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=O. sativa; STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsunoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC  
 RT clone: P0683F02.";  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 RN 12;  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=O. sativa (japonica cultivar-group); STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsunoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC  
 RT clone: OJ1402.H07.";  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP003289; BAB63701.1; -.  
 DR EMBL: AP003415; BAC03323.1; -.  
 DR Gramene: Q94DB8; -.  
 DR GO:0005622; C:intracellular; IEA.  
 DR GO:0003725; F:double-stranded RNA binding; IEA.  
 DR InterPro: IPR001064; Crystalin.  
 DR InterPro: IPR001159; DS\_RBD.

DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
DR PROSITE; PS0137; DS\_RBD; 1.  
SQ SEQUENCE 721 AA; 80021 MW; 6A004326F07646DA CRC64;

Query Match 44.3%; Score 51; DB 10; Length 721;  
Best Local Similarity 44.4%; Pred. No. 9.2;  
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 DIPHPNTHKYLVCESVN 18  
||:|:|:|:|:|:|  
Db 436 DLPYAPDVGDYLVCEDTN 453

Search completed: March 22, 2004, 06:59:11  
Job time : 7.02446 secs





CC humans, other pets, and work or domestic animals. The proteins or  
CC fragments may also be used to diagnose allergies via a skin test. The  
CC proteins and peptides can also be used to raise antibodies, which have a  
CC variety of potential uses. For example, they can be used as vaccines to  
CC passively immunise animals against dust mite hypersensitivity, as  
CC positive controls in test kits and as tools to recover desired dust mite  
CC allergens from a mixture of proteins

XX  
SQ Sequence 20 AA;

Query Match 100.0%; Score 115; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 7.1e-12;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DHPPTNHHKYLVCESVNGG 20  
|||  
1 DHPPTNHHKYLVCESVNGG 20

RESULT 2

AAU96316  
ID AAU96316 standard; peptide; 20 AA.

XX AAU96316;

XX 15-JUL-2002 (first entry)

XX Der HMW-map polypeptide #3.

XX Der HMW-map; American house dust mite; anti-allergic; mite; IGE;

XX mite allergenic protein; immunoglobulin E; hypersensitivity;

XX immunocomplex formation.

XX Dermatophagoides farinae.

XX WO200222807-A2.

XX 21-MAR-2002.

XX 14-SEP-2001; 2001WO-US028730.

XX 14-SEP-2000; 2000US-00662293.

XX (HESK-) HESKA CORP.

XX Mccall CA, Hunter SW, Weber ER;

XX WPI; 2002-351888/38.

XX New mite allergenic protein isolated from Dermatophagoides, designated

XX Der HMW-map protein, useful as a vaccine for treating mite allergy.

XX Claim 12; Page 70; 161pp; English.

CC The invention relates to an isolated mite allergenic protein of  
CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic  
CC acid. The Der HMW-map protein is useful for eliciting an immune response  
CC against Der HMW-map protein. The protein or a reagent comprising a non-  
CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
CC cat) susceptible to or having an allergic response to a mite. A  
CC therapeutic composition is useful for desensitising a host animal to an  
CC allergic response to a mite. The DNA and protein can be used in the  
CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
CC of immunoglobulin (Ig) E or Der HMW-map protein activity associated with a  
CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
CC binding of proteins to IGE, to prevent immunocomplex formation, thus  
CC reducing hypersensitivity responses to mite allergens, and as vaccines  
CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
CC represent Der HMW-map polypeptides of the invention

XX Sequence 20 AA;

Query Match 100.0%; Score 115; DB 5; Length 20;

Best Local Similarity 100.0%; Pred. No. 7.1e-12;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DHPPTNHHKYLVCESVNGG 20  
|||  
1 DHPPTNHHKYLVCESVNGG 20

RESULT 3

AAU96316  
ID AAU96316 standard; peptide; 19 AA.

XX AAU96316;

XX 22-FEB-2000 (first entry)

XX House dust mite allergen protein (map) A/B fragment map (10).

XX Mite allergen protein; map; high molecular weight; HMW-map; allergy;

XX house dust mite; IGE; immunoglobulin E; allergen; mapA; mapB;

XX hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;

XX canine; veterinary; antibody; vaccine; immunisation.

XX Dermatophagoides farinae.

XX WO9954349-A2.

XX 28-OCT-1999.

XX 16-APR-1999; 99WO-US008524.

XX 17-APR-1998; 98US-0062013.

XX 13-MAY-1998; 98US-0085295P.

XX 02-SEP-1998; 98US-0098909P.

XX (HESK-) HESKA CORP.

XX Mccall CA, Hunter SW, Weber ER;

XX WPI; 2000-052700/04.

XX Novel high molecular weight Dermatophagoides nucleic acid polypeptides

XX used to modify an animal's hypersensitivity to mite allergens.

XX Claim 3; Page 70; 154pp; English.

CC Sequences AAU96310-AAU96342 represent proteolytic fragments of  
CC Dermatophagoides farinae high molecular weight mite allergen protein (HMW  
CC -map) composition. The HMW-map composition was isolated from a D. farinae  
CC homogenate by gel filtration, with each fraction being analysed for the  
CC presence of proteins that bound to IGE present in mite-allergic dog  
CC antiserum. The HMW-map composition comprises mapA (a 109 kD protein) and  
CC mapB (98 kD). Mite allergenic proteins and peptides, and nucleic acids  
CC encoding them, may be used in therapeutic compositions to modify an  
CC animal's hypersensitivity reaction to mite allergens. Animals that may be  
CC treated include mammals and birds, especially felines, canines, equines,  
CC humans, other pets, and work or domestic animals. The proteins or  
CC fragments may also be used to diagnose allergies via a skin test. The  
CC proteins and peptides can also be used to raise antibodies, which have a  
CC variety of potential uses. For example, they can be used as vaccines to  
CC passively immunise animals against dust mite hypersensitivity, as  
CC positive controls in test kits and as tools to recover desired dust mite  
CC allergens from a mixture of proteins

XX Sequence 19 AA;

Query Match 94.8%; Score 109; DB 3; Length 19;  
Best Local Similarity 100.0%; Pred. No. 6.7e-11;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DHPPTNHHKYLVCESVNG 19  
|||  
1 DHPPTNHHKYLVCESVNG 19

```

RESULT 4
AAU96324
ID AAU96324 standard; peptide, 19 AA.
XX
XX AAU96324;
XX
XX 15-JUL-2002 (first entry)
XX
XX Der HMW-map polypeptide #11.
XX
XX Der HMW-map; American house dust mite; antiallergic; mite; IgE;
XX mite allergenic protein; immunoglobulin E; hypersensitivity;
XX immunocomplex formation.
XX
XX Dermatophagoides farinae.
XX
XX WO20022807-A2.
XX
XX 21-MAR-2002.
XX
XX 14-SEP-2001; 2001WO-US028730.
XX
XX 14-SEP-2000; 2000US-00662293.
XX
XX (HESK-) HESKA CORP.
XX
XX Mccall CA, Hunter SW, Weber ER;
XX
XX WPI; 2002-351888/38.
XX
XX New mite allergenic protein isolated from Dermatophagoides, designated
XX PT Der HMW-map protein, useful as a vaccine for treating mite allergy.
XX
XX Claim 12; Page 71; 161pp; English.
XX
XX The invention relates to an isolated mite allergenic protein of
XX CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic
XX CC acid. The Der HMW-map protein is useful for eliciting an immune response
XX CC against Der HMW-map protein. The protein or a reagent comprising a non-
XX CC proteinaceous epitope is useful for identifying an animal (e.g., dog,
XX CC cat) susceptible to or having an allergic response to a mite. A
XX CC therapeutic composition is useful for desensitising a host animal to an
XX CC allergic response to a mite. The DNA and protein can be used in the
XX CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition
XX CC of immunoglobulin (Ig)E or Der HMW-map protein activity associated with a
XX CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting
XX CC binding of proteins to IgE, to prevent immunocomplex formation, thus
XX CC reducing hypersensitivity responses to mite allergens, and as vaccines
XX CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342
XX CC represent Der HMW-map polypeptides of the invention.
XX
XX Sequence 19 AA;
XX
XX Query Match 94.8%; Score 109; DB 5; Length 19;
XX Best Local Similarity 100.0%; Pred. No. 6.7e-11;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 DIPHPNTHKYLVCESVNG 19
XX |||||
XX 1 DIPHPNTHKYLVCESVNG 19
XX
XX Db
XX
XX RESULT 5
XX AAU96325
XX ID AAU96325 standard; protein, 536 AA.
XX
XX AAU96325;
XX
XX 22-FEB-2000 (first entry)
XX
XX House dust mite (D. farinae) mite allergen protein (map) pDerf98-536.

```

```

XX
XX Mite allergen protein; map; high molecular weight; HMW-map; allergy;
XX house dust mite; IgE; immunoglobulin E; allergen; map; mapB;
XX hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;
XX canine; veterinary; antibody; vaccine; immunisation.
XX
XX Dermatophagoides farinae.
XX
XX WO9954349-A2.
XX
XX 28-OCT-1999.
XX
XX 16-APR-1999; 99WO-US008524.
XX
XX 17-APR-1998; 98US-00062013.
XX
XX 13-MAY-1998; 98US-0085295P.
XX
XX 02-SEP-1998; 98US-0096909P.
XX
XX (HESK-) HESKA CORP.
XX
XX Mccall CA, Hunter SW, Weber ER;
XX
XX WPI; 2000-052700/04.
XX
XX DR N-PADB; AA238579, AA238580.
XX
XX Novel high molecular weight Dermatophagoides nucleic acid polypeptides
XX PT used to modify an animals' hypersensitivity to mite allergens.
XX
XX Claim 3; Page 125-127; 154pp; English.
XX
XX This sequence represents Dermatophagoides farinae mite allergen protein
XX CC (map) pDerf98-536, the mature form of pDerf98-555 (AAU96323). pDerf98-536
XX CC has a molecular weight of 98 kD, comprising 536 amino acids, and is a
XX CC component of the Dermatophagoides farinae high molecular weight mite
XX CC allergen protein (HMW-map) composition. The HMW-map composition was
XX CC isolated from a D. farinae homogenate by gel filtration, with each
XX CC fraction being analysed for the presence of proteins that bound to IgE
XX CC present in mite-allergic dog antisera. Mite allergenic proteins and
XX CC peptides, and nucleic acids encoding them, may be used in therapeutic
XX CC compositions to modify an animal's hypersensitivity reaction to mite
XX CC allergens. Animals that may be treated include mammals and birds,
XX CC especially felines, canines, equines, humans, other pets, and work or
XX CC domestic animals. The proteins or fragments may also be used to diagnose
XX CC allergies via a skin test. The proteins and peptides can also be used to
XX CC raise antibodies, which have a variety of potential uses. For example,
XX CC they can be used as vaccines to passively immunise animals against dust
XX CC mite hypersensitivity, as positive controls in test kits and as tools to
XX CC recover desired dust mite allergens from a mixture of proteins
XX
XX Sequence 536 AA;
XX
XX Query Match 94.8%; Score 109; DB 3; Length 536;
XX Best Local Similarity 95.0%; Pred. No. 3.3e-09;
XX Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 DIPHPNTHKYLVCESVNG 20
XX |||||
XX 493 DIPHPNTHKYLVCESVNG 512
XX
XX Db
XX
XX RESULT 6
XX AAU96329
XX ID AAU96329 standard; protein, 536 AA.
XX
XX AAU96329;
XX
XX 15-JUL-2002 (first entry)
XX
XX Der HMW-map polypeptide #16.
XX
XX Der HMW-map; American house dust mite; antiallergic; mite; IgE;
XX mite allergenic protein; immunoglobulin E; hypersensitivity;
XX immunocomplex formation.

```

```

XX OS Dermatophagoides farinae.
XX PN MO200222807-A2.
XX PD 21-MAR-2002.
XX PF 14-SEP-2001; 2001WO-US028730.
XX PR 14-SEP-2000; 2000US-00662293.
XX PA (HESK-) HESKA CORP.
XX PI Mccall CA, Hunter SW, Weber ER;
XX DR WPI; 2002-351888/38.
XX DR N-PSDB; ABR69575.
XX PT New mite allergenic protein isolated from Dermatophagoides, designated
XX PT Der HMW-map protein, useful as a vaccine for treating mite allergy.
XX PS Claim 12; Page 125-127; 161pp; English.
XX CC The invention relates to an isolated mite allergenic protein of
XX CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic
XX CC acid. The Der HMW-map protein is useful for eliciting an immune response
XX CC against Der HMW-map protein. The protein or a reagent comprising a non-
XX CC proteiaceous epitope is useful for identifying an animal (e.g., dog,
XX CC cat) susceptible to or having an allergic response to a mite. A
XX CC therapeutic composition is useful for desensitising a host animal to an
XX CC allergic response to a mite. The DNA and protein can be used in the
XX CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition
XX CC of immunoglobulin (Ig) E or Der HMW-map protein activity associated with a
XX CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting
XX CC binding of proteins to IgE, to prevent immunocomplex formation, thus
XX CC reducing hypersensitivity responses to mite allergens, and as vaccines
XX CC against mite allergen hypersensitivity. Sequences AA096314-AA096342
XX CC represent Der HMW-map polypeptides of the invention
XX SQ Sequence 536 AA;

Query Match          94.8%; Score 109; DB 5; Length 536;
Best Local Similarity 95.0%; Pred. No. 3.3e-09;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DIPHPTNIHKYLVCEVNG 20
   |||||
Db 493 DIPHPTNIHKYLVCEVNG 512

RESULT 7
AA096323
ID AA096323 standard; protein; 555 AA.
XX AC AA096323;
XX AC AA096323;
XX DT 22-FEB-2000 (first entry)
XX DE House dust mite (D. farinae) mite allergen protein (map) Pderf98-555.
XX OS Mite allergen protein; map; high molecular weight; HMW-map; allergy;
XX OS house dust mite; IgE; immunoglobulin E; allergen; mapB;
XX OS hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;
XX OS canine; veterinary; antibody; vaccine; immunisation.
XX OS Dermatophagoides farinae.
XX OS Key Location/Qualifiers
XX OS Peptide 1..19 "Signal peptide"
XX OS PT 20..555
XX OS PT Protein /note="Mature Pderf98-555"
XX OS /note="Mature Pderf98-555"

```

```

PN MO09954349-A2.
XX PD 28-OCT-1999.
XX PF 16-APR-1999; 99WO-US008524.
XX PR 17-APR-1998; 98US-00062013.
XX PR 13-MAY-1998; 98US-0085295P.
XX PR 02-SEP-1998; 98US-0098909P.
XX PA (HESK-) HESKA CORP.
XX PI Mccall CA, Hunter SW, Weber ER;
XX DR WPI; 2000-052700/04.
XX DR N-PSDB; AA238575, AA238576, AA238577, AA238578.
XX PT Novel high molecular weight Dermatophagoides nucleic acid polypeptides
XX PT used to modify an animals' hypersensitivity to mite allergens.
XX PS Claim 3; Page 111-113; 154pp; English.
XX CC This sequence represents Dermatophagoides farinae mite allergen protein
XX CC (map) Pderf98-555. Pderf98-555 has a molecular weight of 98 kD,
XX CC comprising 555 amino acids, and is a component of the Dermatophagoides
XX CC farinae high molecular weight mite allergen protein (HMW-map)
XX CC composition. The HMW-map composition was isolated from a D. farinae
XX CC homogenate by gel filtration, with each fraction being analysed for the
XX CC presence of proteins that bound to IgE present in mite-allergic dog
XX CC antisera. Mite allergenic proteins and peptides, and nucleic acids
XX CC encoding them, may be used in therapeutic compositions to modify an
XX CC animal's hypersensitivity reaction to mite allergens. Animals that may be
XX CC treated include mammals and birds, especially felines, canines, equines,
XX CC human, other pets, and work or domestic animals. The proteins or
XX CC fragments may also be used to diagnose allergies via a skin test. The
XX CC proteins and peptides can also be used to raise antibodies, which have a
XX CC variety of potential uses. For example, they can be used as vaccines to
XX CC passively immunise animals against dust mite hypersensitivity, as
XX CC positive controls in test kits and as tools to recover desired dust mite
XX CC allergens from a mixture of proteins
XX SQ Sequence 555 AA;

Query Match          94.8%; Score 109; DB 3; Length 555;
Best Local Similarity 95.0%; Pred. No. 3.5e-09;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DIPHPTNIHKYLVCEVNG 20
   |||||
Db 512 DIPHPTNIHKYLVCEVNG 531

RESULT 8
AA096327
ID AA096327 standard; protein; 555 AA.
XX AC AA096327;
XX AC AA096327;
XX DT 15-JUL-2002 (first entry)
XX DE Der HMW-map polypeptide #14.
XX OS Der HMW-map; American house dust mite; antiallergic; mite; IgE;
XX OS mite allergenic protein; immunoglobulin E; hypersensitivity;
XX OS immunocomplex formation.
XX OS Dermatophagoides farinae.
XX OS Key Location/Qualifiers
XX OS Peptide 1..19 "Signal peptide"
XX OS PT 20..555
XX OS PT Protein /note="Mature Pderf98-555"
XX OS /note="Mature Pderf98-555"
XX OS 14-SEP-2001; 2001WO-US028730.

```

```

XX 14-SEP-2000; 2000US-00662293.
PR (HESK-) HESKA CORP.
XX
PI McCall CA, Hunter SW, Weber ER;
XX
DR WPI; 2002-351888/38.
DR N-PSDB; ABK69571.
XX
PT New mite allergenic protein isolated from Dermatophagoides, designated
PT Der HMW-map protein, useful as a vaccine for treating mite allergy.
XX
PS Claim 12; Page 114-116; 161pp; English.
XX
CC The invention relates to an isolated mite allergenic protein of
CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic
CC acid. The Der HMW-map protein is useful for eliciting an immune response
CC against Der HMW-map protein. The protein or a reagent comprising a non-
CC proteinaceous epitope is useful for identifying an animal (e.g., dog,
CC cat) susceptible to or having an allergic response to a mite. A
CC therapeutic composition is useful for desensitizing a host animal to an
CC allergic response to a mite. The DNA and protein can be used in the
CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition
CC of immunoglobulin (Ig)E or Der HMW-map protein activity associated with a
CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting
CC binding of proteins to IgE, to prevent immunocomplex formation, thus
CC reducing hypersensitivity responses to mite allergens, and as vaccines
CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342
CC represent Der HMW-map polypeptides of the invention
XX
SQ Sequence 555 AA:
XX
Query Match          94.8%; Score 109; DB 5; Length 555;
Best Local Similarity 95.0%; Pred. No. 3.5e-09;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 1 DIPHPNTHKYLVCESVNGG 20
Db 512 DIPHPNTHKYLVCESVNGG 531
XX
RESULT 9
AAU96328
ID AAU96328 standard; protein; 555 AA.
XX
AC AAU96328;
XX
DT 15-JUL-2002 (first entry)
XX
DE Der HMW-map polypeptide #15.
XX
KW Der HMW-map; American house dust mite; antiallergic; mite; IgE;
KW mite allergenic protein; immunoglobulin E; hypersensitivity;
KW immunocomplex formation.
XX
OS Dermatophagoides farinae.
XX
PN MO200222807-A2.
XX
PD 21-MAR-2002.
XX
PF 14-SEP-2001; 2001WO-US028730.
XX
PR 14-SEP-2000; 2000US-00662293.
XX
PA (HESK-) HESKA CORP.
XX
PI McCall CA, Hunter SW, Weber ER;
XX
DR WPI; 2002-351888/38.
DR N-PSDB; ABK69573.
XX

```

```

PT New mite allergenic protein isolated from Dermatophagoides, designated
PT Der HMW-map protein, useful as a vaccine for treating mite allergy.
XX
PS Claim 12; Page 120-122; 161pp; English.
XX
CC The invention relates to an isolated mite allergenic protein of
CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic
CC acid. The Der HMW-map protein is useful for eliciting an immune response
CC against Der HMW-map protein. The protein or a reagent comprising a non-
CC proteinaceous epitope is useful for identifying an animal (e.g., dog,
CC cat) susceptible to or having an allergic response to a mite. A
CC therapeutic composition is useful for desensitizing a host animal to an
CC allergic response to a mite. The DNA and protein can be used in the
CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition
CC of immunoglobulin (Ig)E or Der HMW-map protein activity associated with a
CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting
CC binding of proteins to IgE, to prevent immunocomplex formation, thus
CC reducing hypersensitivity responses to mite allergens, and as vaccines
CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342
CC represent Der HMW-map polypeptides of the invention
XX
SQ Sequence 555 AA:
XX
Query Match          94.8%; Score 109; DB 5; Length 555;
Best Local Similarity 95.0%; Pred. No. 3.5e-09;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 1 DIPHPNTHKYLVCESVNGG 20
Db 512 DIPHPNTHKYLVCESVNGG 531
XX
RESULT 10
AAU52535
ID AAU52535 standard; protein; 490 AA.
XX
AC AAU52535;
XX
DT 06-AUG-2003 (revised)
DT 22-FEB-2000 (first entry)
XX
DE D. pteromyesius 98 kD mite allergen protein (map) PDerp98-490.
XX
KW Mite allergen protein; map; high molecular weight; HMW-map; allergy;
KW house dust mite; IgE; immunoglobulin E; allergen; map;
KW hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;
KW canine; veterinary; antibody; vaccine; immunisation.
XX
OS Dermatophagoides pteromyesius.
XX
FH Key
FH Modified-site 115..117
FH Modified-site /note="Asn is N-glycosylated"
FH Modified-site 240..242
FH /note="Asn is N-glycosylated"
XX
PN MO954349-A2.
XX
PD 28-OCT-1999.
XX
PF 16-APR-1999; 99WO-US008524.
XX
PR 17-APR-1998; 98US-00062013.
PR 13-MAY-1998; 98US-0085295P.
PR 02-SEP-1998; 98US-0098909P.
XX
PA (HESK-) HESKA CORP.
XX
PI McCall CA, Hunter SW, Weber ER;
XX
DR WPI; 2000-052700/04.
DR N-PSDB; AAU38589, AAU38590.
XX

```

PT Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
 used to modify an animals' hypersensitivity to mite allergens.  
 XX  
 XX  
 PS Claim 3; Page 147-149; 154pp; English.  
 CC This sequence represents Dermatophagoides pteronyssinus mite allergen  
 CC protein (map) Pderp98-490, the mature form of Pderp98-509. Pderp98-490  
 CC has a molecular weight of 98 kD, comprising 490 amino acids, and has a  
 CC high degree of homology with the D. farinae mature 98 kD allergen, mapB  
 CC (AA152525). Nucleic acid molecules encoding Pderp98-490 were isolated  
 CC from a D. pteronyssinus cDNA library by hybridisation with a probe  
 CC encoding the D. farinae high molecular weight map (HWM-map) composition.  
 CC Mite allergenic proteins and peptides, and nucleic acids encoding them,  
 CC may be used in therapeutic compositions to modify an animal's  
 CC hypersensitivity reaction to mite allergens. Animals that may be treated  
 CC include mammals and birds, especially felines, canines, equines, humans,  
 CC other pets, and work or domestic animals. The proteins or fragments may  
 CC also be used to diagnose allergies via a skin test. The proteins and  
 CC peptides can also be used to raise antibodies, which have a variety of  
 CC potential uses. For example, they can be used as vaccines to passively  
 CC immunise animals against dust mite hypersensitivity, as positive controls  
 CC in test kits and as tools to recover desired dust mite allergens from a  
 CC mixture of proteins. (Updated on 06-AUG-2003 to correct OS field.)  
 CC  
 SQ Sequence 490 AA;  
 Query Match 71.7%; Score 82.5; DB 3; Length 490;  
 Best Local Similarity 63.6%; Pred. No. 7.5e-05;  
 Matches 14; Conservative 4; Mismatches 1; Indels 3; Gaps 1;  
 QY 2 IPHPTNIHKYLVCESV---NGG 20  
 :||||:|||||:|||||  
 DB 445 LPHTDVHKYLVCEYIATPNGG 466  
 RESULT 11  
 AAU96339  
 ID AAU96339 standard; protein; 490 AA.  
 XX  
 AC AAU96339;  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX  
 DE Der HWM-map polypeptide #26.  
 XX  
 KW Der HWM-map; American house dust mite; anti-allergic; mite; IGE;  
 KW mite allergenic protein; immunoglobulin E; hypersensitivity;  
 KW immunocomplex formation.  
 XX  
 OS Dermatophagoides farinae.  
 XX  
 PN WO200222807-A2.  
 PD 21-MAR-2002.  
 XX  
 PF 14-SEP-2001; 2001WO-US028730.  
 XX  
 PR 14-SEP-2000; 2000US-00662293.  
 XX  
 PA (HESK-) HESKA CORP.  
 XX  
 PI McCall CA, Hunter SW, Weber ER;  
 XX  
 DR WPI; 2002-351888/38.  
 DR N-PSDB; AAK69585.  
 XX  
 PT New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HWM-map protein, useful as a vaccine for treating mite allergy.  
 XX  
 PS Claim 12; Page 144-146; 161pp; English.  
 CC The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HWM-map protein, and its related nucleic

CC acid. The Der HWM-map protein is useful for eliciting an immune response  
 CC against Der HWM-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitising a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HWM-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (IgE or Der HWM-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HWM-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96339-AAU96342  
 CC represent Der HWM-map polypeptides of the invention  
 CC  
 SQ Sequence 490 AA;  
 Query Match 71.7%; Score 82.5; DB 5; Length 490;  
 Best Local Similarity 63.6%; Pred. No. 7.5e-05;  
 Matches 14; Conservative 4; Mismatches 1; Indels 3; Gaps 1;  
 QY 2 IPHPTNIHKYLVCESV---NGG 20  
 :||||:|||||:|||||  
 DB 445 LPHTDVHKYLVCEYIATPNGG 466  
 RESULT 12  
 AA52533  
 ID AA52533 standard; protein; 509 AA.  
 XX  
 AC AA52533;  
 XX  
 DT 06-AUG-2003 (revised)  
 DT 22-FEB-2000 (first entry)  
 XX  
 DE D. pteronyssinus 98 kD mite allergen protein (map) Pderp98-509.  
 XX  
 KW Mite allergen protein; map; high molecular weight; HWM-map; allergy;  
 KW house dust mite; IGE; immunoglobulin E; allergen; mapB;  
 KW hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
 KW canine; veterinary; antibody; vaccine; immunisation.  
 XX  
 OS Dermatophagoides pteronyssinus.  
 XX  
 FE Key Location/Qualifiers  
 FE Peptide 1..19  
 FT /note="Signal peptide"  
 FT Protein 20..509  
 FT /note="Mature Pderp98-509"  
 XX  
 PN WO954349-A2.  
 PD 28-OCT-1999.  
 XX  
 PF 16-APR-1999; 99WO-US008524.  
 XX  
 PR 17-APR-1998; 98US-00062013.  
 PR 13-MAY-1998; 98US-0085285P.  
 PR 02-SEP-1998; 98US-0098909P.  
 XX  
 PA (HESK-) HESKA CORP.  
 XX  
 PI McCall CA, Hunter SW, Weber ER;  
 XX  
 DR WPI; 2000-052700/04.  
 DR N-PSDB; AA238585, AA238586, AA238587, AA238588.  
 XX  
 PT Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
 PT used to modify an animals' hypersensitivity to mite allergens.  
 XX  
 PS Claim 3; Page 134-136; 154pp; English.  
 CC This sequence represents Dermatophagoides pteronyssinus mite allergen  
 CC protein (map) Pderp98-509. Pderp98-509 has a molecular weight of 98 kD,

comprising 509 amino acids, and has a high degree of homology with the D. farinae 98 kD allergen, mapB (AA952523). Nucleic acid molecules encoding CC pDerp98-509 were isolated from a D. pteronyssinus cDNA library by CC hybridisation with a probe encoding the D. farinae high molecular weight CC map (Hmw-map) composition. Mite allergenic proteins and peptides, and CC nucleic acids encoding them, may be used in therapeutic compositions to CC modify an animal's hypersensitivity reaction to mite allergens. Animals CC that may be treated include mammals and birds, especially felines, CC canines, equines, humans, other pets, and work or domestic animals. The CC proteins or fragments may also be used to diagnose allergies via a skin CC test. The proteins and peptides can also be used to raise antibodies, CC which have a variety of potential uses. For example, they can be used as CC vaccines to passively immunise animals against dust mite CC hypersensitivity, as positive controls in test kits and as tools to CC recover desired dust mite allergens from a mixture of proteins. (Updated CC on 06-AUG-2003 to correct OS field.)

SQ Sequence 509 AA;

Query Match 71.7%; Score 82.5; DB 3; Length 509;  
Best Local Similarity 63.6%; Pred. No. 7.9e-05;  
Matches 14; Conservative 4; Mismatches 1; Indels 3; Gaps 1;

QY 2 IPHPTNIRKYLVCESV---NGS 20

Db 464 LPHPTDVHKYLVCEYIATPNGG 485

RESULT 13

AAU96337

AAU96337 standard; protein; 509 AA.

AC AAU96337;

DT 15-JUL-2002 (first entry)

DE Der Hmw-map polypeptide #24.

KM Der Hmw-map; American house dust mite; antiallergic; mite; IgE;

KM mite allergenic protein; immunoglobulin E; hypersensitivity;

OS Immunocomplex formation.

OS Dermatophagoides farinae.

PN WC200222807-A2.

PD 21-MAR-2002.

PF 14-SEP-2001; 2001WO-US028730.

PR 14-SEP-2000; 2000US-00662293.

PA (HESK-) HESKA CORP.

PI Mccall CA, Hunter SM, Weber ER;

DR WPI; 2002-351888/38.

DR N-PSDB; ABR659581.

PT New mite allergenic protein isolated from Dermatophagoides, designated

PT Der Hmw-map protein, useful as a vaccine for treating mite allergy.

PS Claim 12; Page 134-136; 161pp; English.

The invention relates to an isolated mite allergenic protein of CC Dermatophagoides, designated Der Hmw-map protein, and its related nucleic CC acid. The Der Hmw-map protein is useful for eliciting an immune response CC against Der Hmw-map protein. The protein or a reagent comprising a non- CC proteinaceous epitope is useful for identifying an animal (e.g., dog, CC cat) susceptible to or having an allergic response to a mite. A CC therapeutic composition is useful for desensitising a host animal to an CC allergic response to a mite. The DNA and protein can be used in the CC detection of anti-Der Hmw-map antibodies in animal fluids, and inhibition

CC of immunoglobulin (Ig)E or Der Hmw-map protein activity associated with a CC disease. Antibodies that bind to Der Hmw-map are useful for inhibiting CC binding of proteins to IgE, to prevent immunocomplex formation, thus CC reducing hypersensitivity responses to mite allergens, and as vaccines CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342 CC represent Der Hmw-map polypeptides of the invention

SQ Sequence 509 AA;

Query Match 71.7%; Score 82.5; DB 5; Length 509;  
Best Local Similarity 63.6%; Pred. No. 7.9e-05;  
Matches 14; Conservative 4; Mismatches 1; Indels 3; Gaps 1;

QY 2 IPHPTNIRKYLVCESV---NGS 20

Db 464 LPHPTDVHKYLVCEYIATPNGG 485

RESULT 14

AAU96338

AAU96338 standard; protein; 509 AA.

AC AAU96338;

DT 15-JUL-2002 (first entry)

DE Der Hmw-map polypeptide #25.

KM Der Hmw-map; American house dust mite; antiallergic; mite; IgE;

KM mite allergenic protein; immunoglobulin E; hypersensitivity;

OS Immunocomplex formation.

OS Dermatophagoides farinae.

PN WC200222807-A2.

PD 21-MAR-2002.

PF 14-SEP-2001; 2001WO-US028730.

PR 14-SEP-2000; 2000US-00662293.

PA (HESK-) HESKA CORP.

PI Mccall CA, Hunter SM, Weber ER;

DR WPI; 2002-351888/38.

DR N-PSDB; ABR659583.

PT New mite allergenic protein isolated from Dermatophagoides, designated

PT Der Hmw-map protein, useful as a vaccine for treating mite allergy.

PS Claim 12; Page 139-141; 161pp; English.

The invention relates to an isolated mite allergenic protein of CC Dermatophagoides, designated Der Hmw-map protein, and its related nucleic CC acid. The Der Hmw-map protein is useful for eliciting an immune response CC against Der Hmw-map protein. The protein or a reagent comprising a non- CC proteinaceous epitope is useful for identifying an animal (e.g., dog, CC cat) susceptible to or having an allergic response to a mite. A CC therapeutic composition is useful for desensitising a host animal to an CC allergic response to a mite. The DNA and protein can be used in the CC detection of anti-Der Hmw-map antibodies in animal fluids, and inhibition CC of immunoglobulin (Ig)E or Der Hmw-map protein activity associated with a CC disease. Antibodies that bind to Der Hmw-map are useful for inhibiting CC binding of proteins to IgE, to prevent immunocomplex formation, thus CC reducing hypersensitivity responses to mite allergens, and as vaccines CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342 CC represent Der Hmw-map polypeptides of the invention

Query Match 71.7%; Score 82.5; DB 5; Length 509;

Best Local Similarity 63.6%; Pred. No. 7.9e-05;  
Matches 14; Conservative 4; Mismatches 1; Indels 3; Gaps 1;

QY 2 IPHPTNIHKYLCESV--NGG 20  
:|||||:|||||:|||||  
Db 464 LPHPTDVHKYLCCEVATPENG 485

## RESULT 15

AA017674

ID AA017674 standard; protein; 143 AA.

XX AA017674;

AC 05-AUG-2002 (first entry)

XX 05-AUG-2002 (first entry)

DE B tropicalis allergen variant Biot 12(S) protein SEQ ID NO: 19.

XX B tropicalis allergen variant Biot 12(S) protein SEQ ID NO: 19.

XX Mite; allergen; Biot 3; Biot 12(S); Biot 12; dust mite; immunisation;

KW anti-allergic; anti-inflammatory; immunomodulatory; storage mite; asthma;

XX atopic dermatitis; rhinitis; gene therapy; vaccine.

XX atopic dermatitis; rhinitis; gene therapy; vaccine.

OS Blomia tropicalis.

XX Blomia tropicalis.

PN WO200230968-A1.

XX WO200230968-A1.

PD 18-APR-2002.

XX 18-APR-2002.

PF 12-OCT-2001; 2001WO-AU001286.

XX 12-OCT-2001; 2001WO-AU001286.

PR 13-OCT-2000; 2000AU-00000730.

XX 13-OCT-2000; 2000AU-00000730.

PA (UYSI-) UNTV SINGAPORE NAT.

XX (UYSI-) UNTV SINGAPORE NAT.

PI Chua KY, Nge C, Lee BW;

XX Chua KY, Nge C, Lee BW;

DR WPI; 2002-435438/46.

XX WPI; 2002-435438/46.

DR N-PSDB; AAL46687.

XX N-PSDB; AAL46687.

PT New protein allergens from the mite Blomia tropicalis, useful as vaccine

XX New protein allergens from the mite Blomia tropicalis, useful as vaccine

PT for preventing, reducing or ameliorating a B. tropicalis hypersensitivity

XX for preventing, reducing or ameliorating a B. tropicalis hypersensitivity

PT or allergic conditions, e.g. asthma, atopic dermatitis or rhinitis.

XX or allergic conditions, e.g. asthma, atopic dermatitis or rhinitis.

PS Claim 15; Page 133; 150pp; English.

XX Claim 15; Page 133; 150pp; English.

CC The present invention relates to isolated protein allergens from the

XX The present invention relates to isolated protein allergens from the

CC Blomia tropicalis mite, designated Biot 3 and Biot 12(S). The protein

XX Blomia tropicalis mite, designated Biot 3 and Biot 12(S). The protein

CC allergens are useful for preventing, reducing or ameliorating a B.

XX allergens are useful for preventing, reducing or ameliorating a B.

CC tropicalis hypersensitivity condition, such as asthma, atopic dermatitis

XX tropicalis hypersensitivity condition, such as asthma, atopic dermatitis

CC or rhinitis. The present sequence is a protein described in the invention

XX or rhinitis. The present sequence is a protein described in the invention

SQ Sequence 143 AA;

XX Sequence 143 AA;

Query Match 58.3%; Score 67; DB 5; Length 143;

XX Query Match 58.3%; Score 67; DB 5; Length 143;

Best Local Similarity 69.2%; Pred. No. 0.0067; 1; Indels 0; Gaps 0;

XX Best Local Similarity 69.2%; Pred. No. 0.0067; 1; Indels 0; Gaps 0;

Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

XX Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 IPHPTNIHKYLC 14

XX 2 IPHPTNIHKYLC 14

Db 100 IPHPTNVHKYLC 112

XX 100 IPHPTNVHKYLC 112

Search completed: March 22, 2004, 06:51:34  
Job time: 8.08116 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:59:34 ; Search time 4.0916 Seconds  
(without alignments)  
1265.926 Million cell updates/sec

Title: US-09-662-293-3  
Perfect score: 115  
Sequence: 1 DHPPTNKHVYCVESVNGG 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1049977 seqs, 258955339 residues

Total number of hits satisfying chosen parameters: 1049977

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	115	100.0	20	US-10-218-743-3	Sequence 3, Appl1
2	109	94.8	19	US-10-218-743-11	Sequence 21, Appl
3	109	94.8	14	US-10-218-743-21	Sequence 15, Appl
4	109	94.8	14	US-10-218-743-15	Sequence 18, Appl
5	109	94.8	14	US-10-218-743-18	Sequence 41, Appl
6	82.5	71.7	490	US-10-218-743-41	Sequence 35, Appl
7	82.5	71.7	509	US-10-218-743-35	Sequence 38, Appl
8	82.5	71.7	509	US-10-218-743-38	Sequence 5584, Ap
9	59	51.3	67	US-09-664-408A-5584	Sequence 272856,
10	52	40.9	60	US-10-424-599-272856	Sequence 260412,
11	47	40.9	291	US-10-424-599-260412	Sequence 17, Appl
12	47	40.9	291	US-10-222-723-17	Sequence 274767,
13	46.5	40.4	410	US-10-424-599-274767	Sequence 232599,
14	46	40.0	95	US-10-424-599-232599	Sequence 260414,
15	46	40.0	156	US-10-424-599-260414	Sequence 260414,

16	46	40.0	171	12	US-10-425-114-69228	Sequence 69228, A
17	46	40.0	271	10	US-09-898-837A-27	Sequence 27, Appl
18	46	40.0	452	8	US-08-841-636A-35	Sequence 35, Appl
19	46	40.0	891	15	US-10-369-493-2533	Sequence 2533, Ap
20	46	40.0	1344	9	US-09-738-626-6888	Sequence 6888, Ap
21	45	39.1	592	14	US-10-288-556-18	Sequence 18, Appl
22	45	39.1	867	16	US-10-389-566-1146	Sequence 1146, Ap
23	45	39.1	2327	12	US-10-016-248-55	Sequence 55, Appl
24	44.5	38.7	503	12	US-10-425-114-65522	Sequence 65522, A
25	44	38.3	484	15	US-10-369-493-2122	Sequence 2122, Ap
26	44	38.3	585	14	US-10-195-072-4	Sequence 4, Appl1
27	44	38.3	585	14	US-10-195-071-4	Sequence 4, Appl1
28	44	38.3	713	16	US-10-354-437-64	Sequence 64, Appl
29	43.5	37.8	382	12	US-10-424-599-144566	Sequence 14566, A
30	43.5	37.8	382	12	US-10-425-114-59123	Sequence 59123, A
31	43.5	37.8	398	12	US-10-425-114-59943	Sequence 59943, A
32	43	37.4	102	12	US-10-283-122A-76414	Sequence 76414, A
33	43	37.4	148	12	US-10-424-599-263786	Sequence 263786,
34	43	37.4	232	14	US-10-133-973-5	Sequence 5,
35	43	37.4	232	15	US-10-370-570-64	Sequence 64, Appl
36	43	37.4	340	12	US-10-425-114-45458	Sequence 45458, A
37	43	37.4	401	12	US-10-425-114-43164	Sequence 43164, A
38	43	37.4	495	12	US-10-282-122A-65562	Sequence 65562, A
39	43	37.4	502	15	US-10-369-493-18401	Sequence 18401, A
40	43	37.4	755	14	US-10-410-681-4	Sequence 4, Appl1
41	43	37.4	897	15	US-10-369-493-18525	Sequence 18525, A
42	42.5	37.0	297	12	US-10-424-599-277795	Sequence 277795,
43	42	36.5	53	12	US-10-424-599-256323	Sequence 256323,
44	42	36.5	61	12	US-10-424-599-219599	Sequence 219599,
45	42	36.5	72	13	US-10-001-887-131	Sequence 131, App

## ALIGNMENTS

RESULT 1  
US-10-218-743-3  
; Sequence 3, Application US/10218743  
; Publication No. US20030096779A1  
GENERAL INFORMATION:  
; APPLICANT: McCall, Catherine A.  
; APPLICANT: Hunter, Shirley Wu  
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: AL-2-C3  
; CURRENT APPLICATION NUMBER: US/10/218,743  
; CURRENT FILING DATE: 2002-08-13  
; PRIOR APPLICATION NUMBER: US/09/292,225  
; PRIOR FILING DATE: 1999-04-15  
; PRIOR APPLICATION NUMBER: 60/098,909  
; PRIOR FILING DATE: 1998-09-02  
; PRIOR APPLICATION NUMBER: 60/085,295  
; PRIOR FILING DATE: 1998-05-13  
; PRIOR APPLICATION NUMBER: 60/098,565  
; PRIOR FILING DATE: 1998-04-17  
; PRIOR APPLICATION NUMBER: 09/062,013  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Dermatophagoides farinae  
US-10-218-743-3  
Query Match 100.0%; Score 115; DB 14; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DHPPTNKHVYCVESVNGG 20  
1 DHPPTNKHVYCVESVNGG 20  
Db



```
RESULT 2
US-10-218-743-11
; Sequence 11, Application US/10218743
; Publication No. US20030096779A1
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/10/218,743
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US/09/292,225
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: 60/098,909
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/085,295
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/098,565
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: 09/062,013
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-10-218-743-11

Query Match          94.8%; Score 109; DB 14; Length 19;
Best Local Similarity 100.0%; Pred. No. 8.3e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DHPPTNHHKYLVCESVNG 19
DB      1 DHPPTNHHKYLVCESVNG 19

RESULT 3
US-10-218-743-21
; Sequence 21, Application US/10218743
; Publication No. US20030096779A1
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/10/218,743
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US/09/292,225
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: 60/098,909
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/085,295
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/098,565
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: 09/062,013
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-10-218-743-21
```

```
Query Match          94.8%; Score 109; DB 14; Length 536;
Best Local Similarity 95.0%; Pred. No. 2.9e-08;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 DHPPTNHHKYLVCESVNG 20
DB      493 DHPPTNHHKYLVCESVNG 512

RESULT 4
US-10-218-743-15
; Sequence 15, Application US/10218743
; Publication No. US20030096779A1
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/10/218,743
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US/09/292,225
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: 60/098,909
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/085,295
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/098,565
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: 09/062,013
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-10-218-743-15

Query Match          94.8%; Score 109; DB 14; Length 555;
Best Local Similarity 95.0%; Pred. No. 3e-08;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 DHPPTNHHKYLVCESVNG 20
DB      512 DHPPTNHHKYLVCESVNG 531

RESULT 5
US-10-218-743-18
; Sequence 18, Application US/10218743
; Publication No. US20030096779A1
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/10/218,743
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US/09/292,225
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: 60/098,909
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/085,295
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/098,565
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: 09/062,013
; NUMBER OF SEQ ID NOS: 49
```

```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-10-218-743-18

```

Query Match	94.8%;	Score 109;	DB 14;	Length 555;
Best Local Similarity	95.0%;	Pred. No. 3e-08;		
Matches 19; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

QY 1 DHPPTNIHKYLVCESVNGG 20  
|||  
Db 512 DIPHTNIHKYLCFVNNGG 531

```

1      RESULT 6
2      US-10-218-743-41
3      ; Sequence 41, Application US/10218743
4      ; Publication No. US20030096779A1
5      ; GENERAL INFORMATION:
6      ; APPLICANT: McCall, Catherine A.
7      ; APPLICANT: Hunter, Shirley Wu
8      ; APPLICANT: Weber, Eric R.
9      ; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
10     ; TITLE OF INVENTION: AND USES THEREOF
11     ; FILE REFERENCE: AL-2-C3
12     ; CURRENT APPLICATION NUMBER: US/10/218,743
13     ; CURRENT FILING DATE: 2002-08-13
14     ; PRIOR APPLICATION NUMBER: US/09/292,225
15     ; PRIOR FILING DATE: 1999-04-15
16     ; PRIOR APPLICATION NUMBER: 60/098,909
17     ; PRIOR FILING DATE: 1998-09-02
18     ; PRIOR APPLICATION NUMBER: 60/085,295
19     ; PRIOR FILING DATE: 1998-05-13
20     ; PRIOR APPLICATION NUMBER: 60/098,565
21     ; PRIOR FILING DATE: 1998-04-17
22     ; PRIOR APPLICATION NUMBER: 09/002,013
23     ; PRIOR FILING DATE: 1998-04-17
24     ; NUMBER OF SEQ ID NOS: 49
25     ; SOFTWARE: PatentIn Ver. 2.0
26     ; SEQ ID NO 41
27     ; LENGTH: 490
28     ; TYPE: PRT
29     ; ORGANISM: Dermatophagoides farinae
30     ; US-10-218-743-41

```

Query Match	71.7%;	Score 82.5;	DB 14;	Length 490;
Best Local Similarity	63.6%;	Pred. No. 0.00031;		
Matches 14; Conservative	4;	Mismatches 1;	Indels 3;	Gaps 1;

```

QY      2  IPHPTNIHKYLVCSV---NGG  20
      :|||:|||||:  |||
Db      445  LPHPTDVHKYLVCEYIATPNGG  466

```

RESULT 7  
 US-10-218-743-35  
 , Sequence 35, Application US/10218743  
 , Publication No. US20030096779A1  
 , GENERAL INFORMATION:  
 , APPLICANT: McCall, Catherine A.  
 , APPLICANT: Hunter, Shirley Wu  
 , APPLICANT: Weber, Eric R.  
 , TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
 , TITLE OF INVENTION: AND USES THEREOF  
 , FILE REFERENCE: AU-2-C3  
 , CURRENT APPLICATION NUMBER: US/10/218,743  
 , CURRENT FILING DATE: 2002-08-13  
 , PRIOR APPLICATION NUMBER: US/09/292,225  
 , PRIOR FILING DATE: 1999-04-15  
 , PRIOR APPLICATION NUMBER: 60/098,909  
 , PRIOR FILING DATE: 1998-09-02

```

1 PRIOR APPLICATION NUMBER: 60/085,299
2 PRIOR FILING DATE: 1998-05-13
3 PRIOR APPLICATION NUMBER: 60/098,566
4 PRIOR FILING DATE: 1998-04-17
5 PRIOR APPLICATION NUMBER: 09/062,013
6 PRIOR FILING DATE: 1998-04-17
7 NUMBER OF SEQ ID NOS: 49
8 SOFTWARE: PatentIn Ver. 2.0
9 SEQ ID NO 35
10 LENGTH: 509
11 TYPE: prt
12 ORGANISM: Dermatophagoides farinae
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
10
```

Query Match	71.7%;	Score 82.5;	DB 14;	Length 509;
Best Local Similarity	63.6%;	Pred. No. 0.00032;		
Matches 14;	Conservative 4;	Mismatches 1;	Indels 3;	Gaps 1

```

QY      2 IPHPTNIHKYLVCESV---NGG 20
      :|||:|||||: |||
Db      464 LPHPDVAHKYLVCEYIATPNGG 485

```

```

RESULT 8
US-10-218-743-38
; Sequence 38, Application US/10218743
; Publication No. US20030096779A1
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/10/218,743
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US/09/292,225
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: 60/098,909
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/085,295
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/098,565
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: 09/062,013
; PRIOR FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
; US-10-218-743-38

```

Query Match	71.7%;	Score 82.5;	DB 14;	Length 509;
Best Local Similarity	63.6%;	Pred. NO. 0.00032;		
Matches 14; Conservative	4;	Mismatches 1;	Indels 3;	Gaps 1;

```

QY      2  IPHPTNIHKYLVCESV--NGG  20
      :|:|:|:|:|:|:|:|:|:|:|:|
Db     464  LPHPDVIHKYLVCEYIATPNGG  485

```

```

RESULT 9
US-09-864-408A-5584
; Sequence: 5584, Application US/09864408A
; Publication No. US20040009474v1
; GENERAL INFORMATION:
; APPLICANT: leach, Martin D.
; APPLICANT: Shinkets, Richard A.
; TITLE OF INVENTION: No. US20040009474v1 Human Polynucleotides and Polypeptides En
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A

```

CURRENT FILING DATE: 2001-05-24  
PRIOR APPLICATION NUMBER: 60/206,690  
PRIOR FILING DATE: 2000-05-24  
NUMBER OF SEQ ID NOS: 9068  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5584  
LENGTH: 67  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-864-408A-5584

Query Match 51.3%; Score 59; DB 11; Length 67;  
Best Local Similarity 56.2%; Pred. No. 0.15;  
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 IPHPTNIHKYLVCEV 17  
DB 7 IPPPANVHGHWVCESI 22

RESULT 10  
US-10-424-599-272856  
Sequence 272856, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 272856  
LENGTH: 60  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_88411C.1.pap  
US-10-424-599-272856

Query Match 45.2%; Score 52; DB 12; Length 60;  
Best Local Similarity 77.8%; Pred. No. 1.6;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 DIPHTNIH 9  
DB 38 DVPHPTNLH 46

RESULT 11  
US-10-424-599-260412

Sequence 260412, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 260412  
LENGTH: 291  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_77175C.1.pap

US-10-424-599-260412

Query Match 40.9%; Score 47; DB 12; Length 291;  
Best Local Similarity 47.1%; Pred. No. 50;  
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 DIPHTNIHKYLVCEV 17  
DB 88 DIPVTDVHESIQCEAV 104

RESULT 12  
US-10-222-723-17  
Sequence 17, Application US/10222723  
Publication No. US20030145351A1  
GENERAL INFORMATION:  
APPLICANT: Stephen M. Allen  
APPLICANT: Saverio C. Falco  
APPLICANT: Dennis Flint  
APPLICANT: Steven Guttridge  
TITLE OF INVENTION: 3-Deoxy-D-Manno-Octulosonic Acid 8-Phosphate Synthases  
FILE REFERENCE: BR1512 USA  
CURRENT APPLICATION NUMBER: US/10/222,723  
CURRENT FILING DATE: 2002-08-16  
PRIOR APPLICATION NUMBER: 60/312,679  
PRIOR FILING DATE: 2001-08-16  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: Microsoft Word Version 7.0A  
SEQ ID NO 17  
LENGTH: 291  
TYPE: PRT  
ORGANISM: Glycine max  
US-10-222-723-17

Query Match 40.9%; Score 47; DB 14; Length 291;  
Best Local Similarity 47.1%; Pred. No. 50;  
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 DIPHTNIHKYLVCEV 17  
DB 88 DIPVTDVHETIQCEAV 104

RESULT 13  
US-10-424-599-274767  
Sequence 274767, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 274767  
LENGTH: 410  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(410)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_90135C.1.pap  
US-10-424-599-274767

Query Match 40.4%; Score 46.5; DB 12; Length 410;  
Best Local Similarity 43.5%; Pred. No. 86;  
Matches 10; Conservative 4; Mismatches 6; Indels 3; Gaps 1;

```

QY      1 DIPHTNIHKLVCE---SVNGG 20
      |||  :: |||  ||:|
Db      180 DIPSPDVSNLYSEDDGSISNG 202

```

```

RESULT 14
US-10-424-599-232599
Sequence 232599, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongxuei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 232599
LENGTH: 95
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_52061C.1.pep
US-10-424-599-232599

```

Query Match	40.0%;	Score 46;	DB 12;	Length 95;
Best Local Similarity	50.0%;	Pred. No. 22;		
Matches	7;	Conservative	4;	Mismatches 3;
				Indels 0;
				Gaps 0;

```
QY      4 HPTNIHKYLVCESV 17
      | | : | | : | | :
Db     33 HITHTHKFVACEST 46
```

```

RESULT 15
US-10-424-599-260414
Sequence 260414, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ. ID NOS: 285684
SEQ ID NO 260414
LENGTH: 156
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_77177C.1.pep
US-10-424-599-260414

```

Query Match	40.0%;	Score 46;	DB 12;	Length 156;
Best Local Similarity	47.1%;	Pred. No. 37;		
Matches	8;	Conservative	5;	Mismatches 4;
				Indels 0;
				Gaps 0;

```
QY      1 DIPHPTNIHKYLVCESV 17
      ||| |::|: : ||:|
Db      88 DIPIVTDVHETIHCEAV 104
```

```
Search completed: March 22, 2004, 07:45:48
Job time : 4.09116 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:42:54 ; Search time 1.61201 Seconds  
(Without alignments)  
640.518 Million cell updates/sec

Title: US-09-662-293-3  
Perfect score: 115  
Sequence: 1 DIPHTNHHKYLVCESVNGG 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/POCUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	115	100.0	20	US-09-292-225-3	Sequence 3, Appl1
2	109	94.8	19	US-09-292-225-11	Sequence 11, Appl1
3	109	94.8	536	US-09-292-225-21	Sequence 21, Appl1
4	109	94.8	555	US-09-292-225-15	Sequence 15, Appl1
5	109	94.8	555	US-09-292-225-18	Sequence 18, Appl1
6	82.5	71.7	490	US-09-292-225-41	Sequence 41, Appl1
7	82.5	71.7	509	US-09-292-225-35	Sequence 35, Appl1
8	82.5	71.7	509	US-09-292-225-38	Sequence 38, Appl1
9	55	47.8	517	US-09-257-825B-21	Sequence 21, Appl1
10	46	40.0	452	US-09-329-350-35	Sequence 35, Appl1
11	45	40.0	520	US-09-257-825B-20	Sequence 20, Appl1
12	45	39.1	592	US-09-933-711B-18	Sequence 18, Appl1
13	44	38.3	231	US-09-402-515A-16	Sequence 16, Appl1
14	44	38.3	585	US-09-930-181-4	Sequence 4, Appl1
15	43.5	37.8	232	US-09-738-413B-11	Sequence 11, Appl1
16	43	37.4	259	US-09-328-352-5038	Sequence 5038, Ap
17	43	37.4	755	US-09-585-173B-4	Sequence 4, Appl1
18	42	36.5	374	US-09-489-039A-7383	Sequence 7383, Ap
19	42	36.5	603	US-09-930-181-17	Sequence 17, Appl1
20	42	36.5	668	US-09-930-181-2	Sequence 2, Appl1
21	42	36.5	1088	US-08-633-768A-1	Sequence 1, Appl1
22	42	36.5	1088	US-08-280-197-1	Sequence 1, Appl1
23	42	36.5	1092	US-09-275-608-3	Sequence 3, Appl1
24	42	36.5	1843	US-09-413-814-50	Sequence 50, Appl1
25	41.5	36.1	159	US-09-621-377B-4	Sequence 4, Appl1
26	41	35.7	76	5459061-2	Patent No. 5459061
27	41	35.7	103	US-09-732-210-1269	Sequence 1269, Ap

28	41	35.7	182	US-08-858-207A-434	Sequence 434, App
29	41	35.7	202	US-08-155-171B-21	Sequence 21, Appl
30	41	35.7	202	US-08-435-998-21	Sequence 21, Appl
31	41	35.7	509	US-10-095-946-8	Sequence 8, Appl1
32	41	35.7	509	US-09-183-959-8	Sequence 8, Appl1
33	41	35.7	509	US-09-347-650-6	Sequence 6, Appl1
34	41	35.7	509	US-09-535-315-8	Sequence 8, Appl1
35	41	35.7	913	US-08-474-067-6	Sequence 6, Appl1
36	41	35.7	913	US-08-474-068A-6	Sequence 6, Appl1
37	41	35.7	913	US-08-472-481-5	Sequence 5, Appl1
38	41	35.7	920	US-08-930-996A-8	Sequence 8, Appl1
39	41	35.7	926	US-09-489-039A-13928	Sequence 13928, A
40	40.5	35.2	757	US-08-434-000A-6	Sequence 6, Appl1
41	40.5	35.2	757	US-09-312-157-6	Sequence 6, Appl1
42	40	34.8	101	US-08-356-180-2	Sequence 2, Appl1
43	40	34.8	357	US-09-188-579-79	Sequence 79, Appl1
44	40	34.8	357	US-09-315-444-79	Sequence 79, Appl1
45	40	34.8	357	US-09-721-362-79	Sequence 79, Appl1

## ALIGNMENTS

```
RESULT 1
US-09-292-225-3
; Sequence 3, Application US/09292225
; Patent No. 6455686
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; EARLIER FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: 60/098,909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085,295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098,565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062,013
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 3
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-09-292-225-3

Query Match      100.0%; Score 115; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.7e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DIPHTNHHKYLVCESVNGG 20
Db      1 DIPHTNHHKYLVCESVNGG 20

RESULT 2
US-09-292-225-11
; Sequence 11, Application US/09292225
; Patent No. 6455686
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
```

;; CURRENT FILING DATE: 1999-04-15  
;; EARLIER APPLICATION NUMBER: 60/098,909  
;; EARLIER FILING DATE: 1998-09-02  
;; EARLIER APPLICATION NUMBER: 60/085,295  
;; EARLIER FILING DATE: 1998-05-13  
;; EARLIER APPLICATION NUMBER: 60/098,565  
;; EARLIER FILING DATE: 1998-04-17  
;; EARLIER APPLICATION NUMBER: 09/062,013  
;; EARLIER FILING DATE: 1998-04-17  
;; NUMBER OF SEQ ID NOS: 49  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO: 11  
;; LENGTH: 19  
;; TYPE: PRT  
;; ORGANISM: Dermatophagoides farinae  
US-09-292-225-11

Query Match 94.8%; Score 109; DB 4; Length 19;  
Best Local Similarity 100.0%; Pred. No. 4,1e-11;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DIPHPTNIHKYLVCEFSVNG 19  
|||  
Db 1 DIPHPTNIHKYLVCEFSVNG 19

RESULT 3  
US-09-292-225-21

;; Sequence 21, Application US/09292225  
;; Patent No. 6455686  
;; GENERAL INFORMATION:  
;; APPLICANT: McCall, Catherine A.  
;; APPLICANT: Hunter, Shirley Wu  
;; APPLICANT: Weber, Eric R.  
;; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
;; FILE REFERENCE: AL-2-C3  
;; CURRENT APPLICATION NUMBER: US/09/292,225  
;; CURRENT FILING DATE: 1999-04-15  
;; EARLIER APPLICATION NUMBER: 60/098,909  
;; EARLIER FILING DATE: 1998-09-02  
;; EARLIER APPLICATION NUMBER: 60/085,295  
;; EARLIER FILING DATE: 1998-05-13  
;; EARLIER APPLICATION NUMBER: 60/098,565  
;; EARLIER FILING DATE: 1998-04-17  
;; EARLIER APPLICATION NUMBER: 09/062,013  
;; EARLIER FILING DATE: 1998-04-17  
;; NUMBER OF SEQ ID NOS: 49  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO: 21  
;; LENGTH: 536  
;; TYPE: PRT  
;; ORGANISM: Dermatophagoides farinae  
US-09-292-225-21

Query Match 94.8%; Score 109; DB 4; Length 536;  
Best Local Similarity 95.0%; Pred. No. 1,6e-09;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DIPHPTNIHKYLVCEFSVNG 20  
|||  
Db 493 DIPHPTNIHKYLVCEFSVNG 512

RESULT 4  
US-09-292-225-15

;; Sequence 15, Application US/09292225  
;; Patent No. 6455686  
;; GENERAL INFORMATION:  
;; APPLICANT: McCall, Catherine A.  
;; APPLICANT: Hunter, Shirley Wu  
;; APPLICANT: Weber, Eric R.  
;; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS

;; TITLE OF INVENTION: AND USES THEREOF  
;; FILE REFERENCE: AL-2-C3  
;; CURRENT APPLICATION NUMBER: US/09/292,225  
;; CURRENT FILING DATE: 1999-04-15  
;; EARLIER APPLICATION NUMBER: 60/098,909  
;; EARLIER FILING DATE: 1998-09-02  
;; EARLIER APPLICATION NUMBER: 60/085,295  
;; EARLIER FILING DATE: 1998-05-13  
;; EARLIER APPLICATION NUMBER: 60/098,565  
;; EARLIER FILING DATE: 1998-04-17  
;; EARLIER APPLICATION NUMBER: 09/062,013  
;; EARLIER FILING DATE: 1998-04-17  
;; NUMBER OF SEQ ID NOS: 49  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO: 15  
;; LENGTH: 555  
;; TYPE: PRT  
;; ORGANISM: Dermatophagoides farinae  
US-09-292-225-15

Query Match 94.8%; Score 109; DB 4; Length 555;  
Best Local Similarity 95.0%; Pred. No. 1,6e-09;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DIPHPTNIHKYLVCEFSVNG 20  
|||  
Db 512 DIPHPTNIHKYLVCEFSVNG 531

RESULT 5  
US-09-292-225-18

;; Sequence 18, Application US/09292225  
;; Patent No. 6455686  
;; GENERAL INFORMATION:  
;; APPLICANT: McCall, Catherine A.  
;; APPLICANT: Hunter, Shirley Wu  
;; APPLICANT: Weber, Eric R.  
;; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
;; FILE REFERENCE: AL-2-C3  
;; CURRENT APPLICATION NUMBER: US/09/292,225  
;; CURRENT FILING DATE: 1999-04-15  
;; EARLIER APPLICATION NUMBER: 60/098,909  
;; EARLIER FILING DATE: 1998-09-02  
;; EARLIER APPLICATION NUMBER: 60/085,295  
;; EARLIER FILING DATE: 1998-05-13  
;; EARLIER APPLICATION NUMBER: 60/098,565  
;; EARLIER FILING DATE: 1998-04-17  
;; EARLIER APPLICATION NUMBER: 09/062,013  
;; EARLIER FILING DATE: 1998-04-17  
;; NUMBER OF SEQ ID NOS: 49  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO: 18  
;; LENGTH: 555  
;; TYPE: PRT  
;; ORGANISM: Dermatophagoides farinae  
US-09-292-225-18

Query Match 94.8%; Score 109; DB 4; Length 555;  
Best Local Similarity 95.0%; Pred. No. 1,6e-09;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DIPHPTNIHKYLVCEFSVNG 20  
|||  
Db 512 DIPHPTNIHKYLVCEFSVNG 531

RESULT 6  
US-09-292-225-41

;; Sequence 41, Application US/09292225  
;; Patent No. 6455686  
;; GENERAL INFORMATION:  
;; APPLICANT: McCall, Catherine A.

```

; APPLICANT: Hunter, Shirley Wu
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; EARLIER FILING DATE: 1998-04-15
; EARLIER APPLICATION NUMBER: 60/098,909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085,295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098,565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062,013
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41
; LENGTH: 490
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
; US-09-292-225-41

Query Match          71.7%; Score 82.5; DB 4; Length 490;
Best Local Similarity 63.6%; Pred. No. 2.7e-05;
Matches 14; Conservative 4; Mismatches 1; Indels 3; Gaps 1;

QY      2 IPHPTNHHKYLVCESV---NGG 20
DB      445 LPHPTDVHKKYLVCEYIATPNGG 466

RESULT 7
US-09-292-225-35
; Sequence 35, Application US/09292225
; Patent No. 6455686
; GENERAL INFORMATION:
; APPLICANT: Hunter, Eric R.
; APPLICANT: Hunter, Shirley Wu
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; EARLIER FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: 60/098,909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085,295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098,565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062,013
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
; US-09-292-225-35

Query Match          71.7%; Score 82.5; DB 4; Length 509;
Best Local Similarity 63.6%; Pred. No. 2.8e-05;
Matches 14; Conservative 4; Mismatches 1; Indels 3; Gaps 1;

QY      2 IPHPTNHHKYLVCESV---NGG 20
DB      464 LPHPTDVHKKYLVCEYIATPNGG 485

RESULT 8
US-09-292-225-38
; Sequence 38, Application US/09292225
```

```

; Patent No. 6455686
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Hunter, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; EARLIER FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: 60/098,909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085,295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098,565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062,013
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
; US-09-292-225-38

Query Match          71.7%; Score 82.5; DB 4; Length 509;
Best Local Similarity 63.6%; Pred. No. 2.8e-05;
Matches 14; Conservative 4; Mismatches 1; Indels 3; Gaps 1;

QY      2 IPHPTNHHKYLVCESV---NGG 20
DB      464 LPHPTDVHKKYLVCEYIATPNGG 485

RESULT 9
US-09-257-825B-21
; Sequence 21, Application US/09257825B
; Patent No. 6403352
; GENERAL INFORMATION:
; APPLICANT: Poovalah, Rachetlita W.
; APPLICANT: Patil, Shameekumar
; APPLICANT: Takezawa, Daijuke
; TITLE OF INVENTION: Compositions and Methods for Production of Male-Sterile Plants
; FILE REFERENCE: 4630-51993
; CURRENT APPLICATION NUMBER: US/09/257,825B
; EARLIER FILING DATE: 1999-02-25
; EARLIER APPLICATION NUMBER: US 08/655,352
; PRIOR FILING DATE: 1996-05-23
; PRIOR APPLICATION NUMBER: US 60/014,743
; PRIOR FILING DATE: 1996-03-28
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 517
; TYPE: PRT
; ORGANISM: Nicotiana tabacum
; US-09-257-825B-21

Query Match          47.8%; Score 55; DB 4; Length 517;
Best Local Similarity 58.8%; Pred. No. 0.8;
Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY      3 PHPTNHHKYLVCESYNG 19
DB      97 PHPNVHLYDVCEDPSG 113

RESULT 10
US-09-329-350-35
; Sequence 35, Application US/09329350
; Patent No. 6184019
; GENERAL INFORMATION:
```

```

1  APPLICANT: Miettinen-Ononen, Arja
2  APPLICANT: Londeshborough, John
3  APPLICANT: Vehmaanger, Jari
4  APPLICANT: Haakana, Heili
5  APPLICANT: M ntyl, Arja
6  APPLICANT: Lantto, Raita
7  APPLICANT: Elovalinto, Minna
8  APPLICANT: Joutsjoki, Vesa
9  APPLICANT: Paloheimo, Marja
10 APPLICANT: Suominen, Pirkko
11 TITLE OF INVENTION: NOVEL CELLULASES, THE GENES ENCODING THEM AND
12 TITLE OF INVENTION: US85 THERMOP
13 NUMBER OF SEQUENCES: 45
14 CORRESPONDENCE ADDRESSES:
15 ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
16 STREET: 1100 New York Avenue, N.W., Suite 600
17 CITY: Washington
18 STATE: D.C.
19 COUNTRY: USA
20 ZIP: 20005
21 COMPUTER READABLE FORM:
22 MEDIUM TYPE: Diskette, 3.50 inch
23 COMPUTER: IBM PC compatible
24 OPERATING SYSTEM: PC-DOS/MS-DOS
25 SOFTWARE: Patent in Release #1.0, Version #1.30 (EPO)
26 CURRENT APPLICATION DATA:
27 APPLICATION NUMBER: US/09/329,350
28 FILING DATE: Herzewich
29 CLASSIFICATION:
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER: US 08/841,636
32 FILING DATE: 30-APR-1997
33 PRIOR APPLICATION DATA:
34 APPLICATION NUMBER: US 60/005,335
35 FILING DATE: 17-OCT-1995
36 PRIOR APPLICATION DATA:
37 APPLICATION NUMBER: US 60/007,926
38 FILING DATE: 04-DEC-1995
39 PRIOR APPLICATION DATA:
40 APPLICATION NUMBER: US 60/020,840
41 FILING DATE: 28-JUN-1996
42 PRIOR APPLICATION DATA:
43 APPLICATION NUMBER: US 08/732,181
44 FILING DATE: 16-OCT-1996
45 PRIOR APPLICATION DATA:
46 APPLICATION NUMBER: PCT/FI96/00550
47 FILING DATE: 17-OCT-1996
48 ATTORNEY/AGENT INFORMATION:
49 NAME: Shea Jr., Timothy
50 REGISTRATION NUMBER: 41,306
51 REFERENCE/DOCKET NUMBER: 1716.0510006/MAC/TUS
52 TELECOMMUNICATION INFORMATION:
53 TELEPHONE: (202)371-2600
54 TELEFAX: (202)371-2540
55 INFORMATION FOR SEQ ID NO: 35:
56 SEQUENCE CHARACTERISTICS:
57 LENGTH: 452 amino acids
58 TYPE: amino acid
59 STRANDEDNESS:
60 TOPOLOGY: linear
61 MOLECULE TYPE: protein
62 ORIGINAL SOURCE:
63 ORGANISM: Melanocarpus albomyces
64 STRAIN: ALKO4237
65 FEATURE:
66 NAME/KEY: Protein
67 LOCATION: 1..452
68 OTHER INFORMATION: /label= 50K-cellulase-B
69
70 US-09-329-350-35

```

QY 3 PHPTNTHKYLVCESVNGG 20  
|| :: || || : ||  
Db 249 PHACTNETHVCEFTNCG 266

RESULT 11  
US-09-257-825B-20

; Sequence 20, Application US/09257825E  
; Patent No. 6403352

; GENERAL INFORMATION:  
: APPLICANT: Poovaiyah

APPLICANT: Patil, Shameekumar  
ADDRESSEE: Maheshwari, Dattatraya

TITLE OF INVENTION: Compositions and Methods for Production of Male-Sterile Plants

FILE REFERENCE: 4630-51993  
CURRENT APPLICATION NUMBER

CURRENT FILING DATE: 1999-02-25  
PRIOR APPLICATION NUMBER: US 08/655 352

PRIOR FILING DATE: 1996-05-23

; PRIOR APPLICATION NUMBER: US 60/014,743  
; PRIOR FILING DATE: 1996-03-28

```

; NUMBER OF SEQ ID NOS: 28
SOFTWARE: PatentIn version 3.1

```

SEQ	ID	NO	20
1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16
17	18	19	20
21	22	23	24
25	26	27	28
29	30	31	32
33	34	35	36
37	38	39	40
41	42	43	44
45	46	47	48
49	50	51	52
53	54	55	56
57	58	59	60
61	62	63	64
65	66	67	68
69	70	71	72
73	74	75	76
77	78	79	80
81	82	83	84
85	86	87	88
89	90	91	92
93	94	95	96
97	98	99	100

TYPE: PRT

ORGANISM: *Lilium longiflorum*;  
US-09-257-825B-20

Query Match 40.0%: Score 46: DB 4: Length 520

Best Local Similarity 52.9%; Pred. No. 23;

Matches	9	conservative	1	mismatches	1	indels	0	gaps	0
---------	---	--------------	---	------------	---	--------	---	------	---

QY 3 PHPTNIHKYLVCSVNG 19

Db 100 PHPNVIHLHDVYEDANG 116

1  
1  
0  
4  
E  
2  
0

RESULT 12  
US-08-933-711B-18

Sequence 18, Appl 11  
Patent No. 6514724

GENERAL INFORMATION:

APPLICANT: Chuang, Pao-Tien

```

; TITLE OF INVENTION: HEDGEHOG INTERACTING P
; FILE REFERENCE: HUV-024.01

```

CURRENT APPLICATION NUMBER: US/08/93  
CURRENT FILING DATE: 1997-09-19

PRIOR APPLICATION NUMBER: 60/

;; PRIOR FILING DATE: 1996-09-  
; NUMBER OF SEQ ID NOS: 18

```

; SOFTWARE: Patentin Ver. 2.1
; SEO ID NO 18

```

LENGTH: 592  
TYPE: DBT

ORGANISM: Artificial Sequence

```

;
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Consensus

```

OTHER INFORMATION: sequence  
US-08-933-711B-18

Current Month	2014	Current AE	DD 4	Forecast EGS
0	0	0	0	0
1	1	1	1	1
2	2	2	2	2
3	3	3	3	3
4	4	4	4	4
5	5	5	5	5
6	6	6	6	6
7	7	7	7	7
8	8	8	8	8
9	9	9	9	9
10	10	10	10	10
11	11	11	11	11
12	12	12	12	12
13	13	13	13	13
14	14	14	14	14
15	15	15	15	15
16	16	16	16	16
17	17	17	17	17
18	18	18	18	18
19	19	19	19	19
20	20	20	20	20
21	21	21	21	21
22	22	22	22	22
23	23	23	23	23
24	24	24	24	24
25	25	25	25	25
26	26	26	26	26
27	27	27	27	27
28	28	28	28	28
29	29	29	29	29
30	30	30	30	30
31	31	31	31	31
32	32	32	32	32
33	33	33	33	33
34	34	34	34	34
35	35	35	35	35
36	36	36	36	36
37	37	37	37	37
38	38	38	38	38
39	39	39	39	39
40	40	40	40	40
41	41	41	41	41
42	42	42	42	42
43	43	43	43	43
44	44	44	44	44
45	45	45	45	45
46	46	46	46	46
47	47	47	47	47
48	48	48	48	48
49	49	49	49	49
50	50	50	50	50
51	51	51	51	51
52	52	52	52	52
53	53	53	53	53
54	54	54	54	54
55	55	55	55	55
56	56	56	56	56
57	57	57	57	57
58	58	58	58	58
59	59	59	59	59
60	60	60	60	60
61	61	61	61	61
62	62	62	62	62
63	63	63	63	63
64	64	64	64	64
65	65	65	65	65
66	66	66	66	66
67	67	67	67	67
68	68	68	68	68
69	69	69	69	69
70	70	70	70	70
71	71	71	71	71
72	72	72	72	72
73	73	73	73	73
74	74	74	74	74
75	75	75	75	75
76	76	76	76	76
77	77	77	77	77
78	78	78	78	78
79	79	79	79	79
80	80	80	80	80
81	81	81	81	81
82	82	82	82	82
83	83	83	83	83
84	84	84	84	84
85	85	85	85	85
86	86	86	86	

Query match	score	DB #	dengchi
Best Local Similarity	38.9%	Pred. No. 38;	

Matches 7; Conservative 4; Mismatches 7; Indels 0; Gaps 0

QY 2 IPHP7NIHKYLVCSVNG 19

Db 356 VDHDPDNLTLICSDSNG 373

### RESULT 13





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:40:28 ; Search time 1.30072 Seconds  
(without alignments)  
1479.047 Million cell updates/sec

Title: US-09-662-293-4  
Perfect score: 105  
Sequence: 1 DPAKMSPPGFTVGESEGLS 20

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78: \*  
1: p1r1: \*  
2: p1r2: \*  
3: p1r3: \*  
4: p1r4: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	47.6	408	T13387	hypothetical prote
2	47	44.8	573	A86287	hypothetical prote
3	46	43.8	433	C87552	aspartate transcar
4	45	42.9	141	S09804	hypothetical prote
5	44	41.9	111	H72501	hypothetical prote
6	44	41.9	408	A99238	conserved hypotet
7	44	41.9	410	H72502	hypothetical prote
8	44	41.9	690	F70628	probable pta prote
9	44	41.9	813	T02672	hypothetical prote
10	44	41.9	835	I55603	reduced hepatic gl
11	44	41.9	1539	S65775	homeotic protein H
12	44	41.9	3224	S58884	Ran-binding protei
13	43	41.0	291	F81444	probable integral
14	43	41.0	305	AG2399	hypothetical prote
15	43	41.0	359	H70876	hypothetical prote
16	43	41.0	367	JC6087	helix-loop-helix t
17	43	41.0	394	E64028	hypothetical prote
18	43	41.0	663	AE3290	protonate-CoA lig
19	43	41.0	728	TS1071	related to trfA pr
20	43	41.0	817	S53919	hypothetical prote
21	43	41.0	883	B95237	alcohol dehydrogen
22	43	41.0	890	C98101	alcohol-acetaldehy
23	43	41.0	891	AE0358	clp ATPase (import
24	42.5	40.5	54463	R0503	10 protein - Rhl
25	42.5	40.5	561	S72618	hypX protein - Rhl
26	42.5	40.5	864	F98196	probable clpA/B-ty
27	42.5	40.5	867	AH0437	clp ATPase (import
28	42.5	40.5	869	G82499	clpB protein VCA01
29	42.5	40.5	892	AD3050	ATP-dependent Clp

30	42.5	40.5	923	2	G90656	hypothetical prote
31	42.5	40.5	923	2	G85507	probable proteinas
32	42.5	40.5	3176	2	CGH134	collagen alpha 3(V
33	42	40.0	190	2	A87263	hypothetical prote
34	42	40.0	231	2	A82164	hypothetical prote
35	42	40.0	269	2	H84378	hypothetical prote
36	42	40.0	438	1	MMAD40	late 100K protein
37	42	40.0	623	2	A49112	sodium-glucose cot
38	42	40.0	677	2	I49045	probable ATPase SK
39	42	40.0	885	2	A84373	leucine-cRNA synth
40	42	40.0	889	2	JC5576	inter-alpha-trypsi
41	42	40.0	907	1	A57429	aldehyde oxidase (
42	41.5	39.5	80	2	AF2125	hypothetical prote
43	41.5	39.5	224	2	G87377	hypothetical prote
44	41	39.0	184	2	AD0314	probable membrane
45	41	39.0	198	2	AC3305	hypothetical prote

## ALIGNMENTS

RESULT 1  
T13387  
hypothetical protein 115C2.8 - fruit fly (Drosophila melanogaster)  
C/Species: Drosophila melanogaster  
C/Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 19-Jan-2001  
C/Accession: T13387  
R/Salles, C.; Valenti, P.; Darlameisou, A.; Henderson, N.; Campbell, L.; Glover, D.  
submitted to the EMBL Data Library, May 1999  
A/Description: Sequencing the distal X chromosome of Drosophila melanogaster.  
A/Reference number: Z17655  
A/Accession: T13387  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-408 <CAT>  
A/Cross-references: EMBL:AL031561; NID:e1320978; PID:e1320993; PIDN:CAA20887.1  
C/Genetic: 8  
A/Cross-references: FlyBase:FBgn0020381  
A/Introns: 340/1  
A/Note: BG:115C2.8  
C/Superfamily: Drosophila melanogaster hypothetical protein 115C2.8

Query Match 47.6%; Score 50; DB 2; Length 408;  
Best Local Similarity 40.0%; Pred. No. 4.7;  
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 DPAKMSPPGFTVGESEGLS 20  
DB 327 DQEGFDPTVIGPDESL 346

RESULT 2  
A86287  
hypothetical protein F9L1.22 - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001  
C/Accession: A86287  
R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federici, N.A.; Kaul, S.; White, O.; Alonso  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.  
arsen, N.F.; Hughes, B.; Hulst, L.  
Nature 408, 816-820, 2000  
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A/Rifle: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A/Reference number: A86141; MUID:21016719; PMID:11130712  
A/Accession: A86287  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-573 <STO>  
A/Cross-references: GB:AE005172; NID:gs103826; PIDN:AA039656.1; GSPDB:GNO0141

C:Genetics:  
A:Map position: 1

Query Match 44.8%; Score 47; DB 2; Length 573;  
Best Local Similarity 47.1%; Pred. No. 21;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 PAKGSPPGFIVGEGV 18  
DB 477 PTFGWAYPGYQSEDDV 493

RESULT 3

C87552  
aspartate transcarbamoylase, pyroC' subunit [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001

C:Accession: C87552  
R:Nilnerman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Esmolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98: 4136-4141, 2001  
A:Title: Complete genome sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: C87552

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-433 <STO>

A:Cross-references: GB:AE006673; NID:gl3423989; PIDN:AAK24415.1; GSPDB:GN00148

A:Gene: CC2444

C:Superfamily: Bacillus dihydroorotase; Bacillus dihydroorotase homology

Query Match 43.8%; Score 46; DB 2; Length 433;  
Best Local Similarity 40.0%; Pred. No. 22;  
Matches 8; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 DPAKMSPPGFIVGEGV 20  
DB 16 DPESGYDPGGVIVSEGVIT 35

RESULT 4

S09804  
hypothetical protein UL41 - human cytomegalovirus (strain AD169)

C:Species: human cytomegalovirus, human herpesvirus 5

A:Note: host Homo sapiens (man)

C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 28-Jul-2000

C:Accession: S09804

R:Chen, M.S.; Bankier, A.T.; Beck, S.; Bohni, R.; Brown, C.M.; Cerny, R.; Horenell, T.;

M.; Barrell, B.G.

Curr. Top. Microbiol. Immunol. 154, 125-169, 1990

A:Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus

A:Reference number: S09749; MUID:90269039; PMID:2161319

A:Accession: S09804

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-141 <CHE>

A:Cross-references: EMBL:X17403; NID:959591; PIDN:CAA35400.1; PID:e27240; PID:G1813967

A:Note: this sequence was submitted to the EMBL Data Library, December 1989

C:Superfamily: human cytomegalovirus hypothetical protein UL41

Query Match 42.9%; Score 45; DB 2; Length 141;  
Best Local Similarity 53.3%; Pred. No. 8.7;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 AKGMSPPGFIVGEGV 17  
DB 76 AKDLPPGGYRGVGRG 90

RESULT 5

H72501  
hypothetical protein APE1990 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix

C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jun-2000

C:Accession: H72501

R:Kawabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takah

awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K

DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr

A:Reference number: A72450; MUID:99310339; PMID:10382966

A:Accession: H72501

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-111 <KAW>

A:Cross-references: DDBJ:AB000663; NID:95105654; PIDN:BAAB1000.1; PID:d1044786; PID:9510

A:Experimental source: strain K1

C:Genetics:

A:Gene: APE1990

C:Superfamily: Aeropyrum pernix hypothetical protein APE1990

Query Match 41.9%; Score 44; DB 2; Length 111;  
Best Local Similarity 44.4%; Pred. No. 9.6;  
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 DPAKMSPPGFIVGEGV 18  
DB 85 EAARGVGPAGFPVDDGV 102

RESULT 6

A99238  
conserved hypothetical protein [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 15-Jun-2001

C:Accession: A99238

R:Jong, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-

g, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P

arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.

submitted to Genbank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: A99238

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-408 <KJR>

A:Cross-references: GB:AE006641; NID:gl3814052; PIDN:AAK41160.1; GSPDB:GN00155

C:Genetics:

A:Gene: SSO0872

C:Superfamily: hypothetical protein AF1590

Query Match 41.9%; Score 44; DB 2; Length 408;  
Best Local Similarity 57.1%; Pred. No. 42;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 KGMSPPGFIVGEGV 17  
DB 50 EGSTPPSGIVGK 63

RESULT 7

H72502  
hypothetical protein APE1990 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix

C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000

C:Accession: H72502

R:Kawabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takat

awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K

DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr

A:Reference number: A72450; MUID:99310339; PMID:10382966

A:Accession: H72502

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-410 <XAM>  
A:Cross-references: DDBA:AP000063; NID:G5105654; PIDN:BAAB1008.1; PID:G51056969  
A:Experimental source: strain K1  
C:GeneticCS:  
A:Gene: APR1998  
C:Superfamily: hypothetical protein AF1590

Query Match	41.9%;	Score 44;	DB 2;	Length 410;
Best Local	Similarity	50.0%;	Pred. No. 42;	
Matches	7;	Conservative	3;	Mismatches 4; Indels 0; Gaps 0;

```

QY      4 KGMSPPGFIVGEEG 17
         :|:|:|:|
Db      47 QGSTPPSILVGERG 60

```

RESULT 8  
F70628  
probable pta protein - Mycobacterium tuberculosis (strain H37Rv)  
C. Species: Mycobacterium tuberculosis

CitDate: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #ext\_change 20-Jun-2000  
C/Accession: F70628  
R/Colo, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
R/Conor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentile, S.; Hamlin, N.; Holroyd, S.  
R/landrem, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
NAutheors: Squares, R., Stulston, J.E., Taylor, K., Whitehead, S., Barrell, B.G.  
ArTitle: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
ArReference number:AV0500, MUID:98295987, PMID:9634230

Query Match	41.9%	Score 44	DB 2	Length 690
Best Local	Similarity 40.0%	Pred. No. 76		
Matches 8	Conservative 5	Mismatches 7	Indels 0	Gaps 0

```
QY      1 DPAKGMSPPGFIVGEEGVL 20
        | : : : : :
D5      182 DALRRFTPPSYVVPPEPL 201
```

RESULT 9  
T02672  
hypothetical protein R31449\_3 - human (fragment)  
D:\Genetics\Homo sapiens (homo)

C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: T02672  
R:Blamerdin, J.E.; McGready, P.M.; Skowronski, E.; Adamson, A.W.; Burkhart-Schultz, K.; C  
J.; Dangsan, L.; Poundstone, P.; Christensen, M.; Georgescu, A.; Avila, U.; Liu, S.; Alai,  
P.; Quan, G.; Krommiller, E.; Kellano, A.; Montgomeri, M.; Ow, D.; Nolan, M.  
submitted to the EMBL Data Library, June 1998  
A:Authors: Strong, S.; Kobayashi, A.; Olsen, A.S.; Carrano, A.V.  
A:Description: Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a serin  
A:Reference number: Z14696

Query Match	41.9%;	Score 44;	DB 2;	Length 813;
Best Local Similarity	53.8%;	Pred. No. 92;		

Matches	7,	Conservative	2,	Mismatches	4,	Indels	0,	Gaps	0;
Qy	2	PAKMSPPGFIVG	14						
		:	:						
Db	675	PAAGIPPGGLTG	687						

RESULT 10  
TEEC03

reduced heptad:glutathione transporter with canalicular features - rat  
 C.Species: Rattus norvegicus (Norway rat)  
 C.Date: 26-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 07-Feb-1997  
 C.Accession: I55603  
 R.Yi, J.R.; Lu, S.; Fernandez-Checa, J.; Kaplowitz, N.  
 J.Clin. Invest. 93, 1841-1845, 1994  
 A.Title: Expression cloning of a rat hepatic reduced glutathione transporter with canalicular features  
 A.Reference number: I55603; MUID:94216535; PMID:1616683  
 A.Accession: I55603  
 A.Status: preliminary; translated from GB/EMBL/DBJ  
 A.Molecule type: mRNA  
 A.Residues: 1-835 <RES>  
 A.Cross-references: EMBL:U06645; NID:g501037; PID:g501038  
 C.Keywords: liver

Query Match	41.9%	Score 44	DB 2	Length 835
Best Local Similarity	64.3%	Pred. No.	95	
Matches	9	Conservative	0	Mismatches 5
				Indels 0
				Gaps 0

QY 2 PAKGMSPPGFIVGE 15  
Db 727 PTKGALSPPFIVGE 740

## RESULT 11

homeotic protein Hox2b - maize  
C:Species: Zea mays (maize)  
C:Date: 14-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 11-Jan-2002  
C:Accession: S65775; S72455  
R:Killing, B.; Ueberlacker, B.; Korthage, C.; Werr, W.  
Plant Mol. Biol. 30, 439-453, 1996  
A>Title: Zmhox: a novel class of maize homeobox genes.  
A:Reference number: S65774; MUID:96189260; PMID:8605257  
A:Accession: S65775  
A>Status: preliminary; nucleic acid sequence not shown  
A:Residues: 1-1539 <KLI>  
A:Molecule type: mRNA  
A:Cross-references: EMBL:X89761  
R:Werr, W.  
submitted to the EMBL Data Library, July 1995  
A:Reference number: S72455  
A:Accession: S72455  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-379,'S',381-1539 <MER>  
A:Cross-references: EMBL:X89761; NID:G1143706; PIDN:CAA61910.1; PID:G1143707  
C:Genetics:  
A:Gene: Hox2b  
C:Superfamily: unassigned homeobox proteins; homeobox homology  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:1804-858/Domain: homeobox homology <HOK1>  
F:1040-1096/Domain: homeobox homology <HOK2>

Query Match	41.9%	Score 44	DB 2	Length 1539
Best Local Similarity	53.8%	Pred. No. 1.9e+02		
Matches	7	Conservative	3	Mismatches 3
				Indels 0
				Gaps 0
QY	3	AKGMSPPGFIQGE	15	
	: :	:	:	
Db	1217	ARGVGPPIYTPGB	1229	

## RESULT 12

S58884

Ran-binding protein 2 - human

N:Alternate names: giant nucleopore protein Nup358; nucleoporin Nup358; RanBP2 protein

C:Species: Homo sapiens (man)

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000

C:Accession: S58884; A57545

R:Yokoyama, N.; Hayashi, N.; Seki, T.; Panse, N.; Ohba, T.; Nishii, K.; Kuma, K.; Hayash

Nature 376, 184-188, 1995

A:Title: A giant nucleopore protein that binds Ran/TC4.

A:Reference number: S58884; PMID:95327194; PMID:7603572

A:Accession: S58884

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-3224 &lt;YOK&gt;

A:Cross-references: EMBL:D42063; NID:G924266; PIDN:BA07662.1; PID:gl009337

A:Experimental source: cell type B-lymphocyte

R:Mu, J.; Matunis, M.J.; Kraemer, D.; Blobel, G.; Coutavas, E.

J. Biol. Chem. 270, 14209-14213, 1995

A:Title: Nup358, a cytoplasmically exposed nucleoporin with peptide repeats, Ran-GTP bit

A:Reference number: A57545; PMID:95294031; PMID:775481

A:Accession: A57545

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-776, 'R', 778-783, 'R', 785-3224 &lt;WTA&gt;

A:Cross-references: GB:141840; NID:9857367; PIDN:AA041758.1; PID:9857368

A:Experimental source: cell line HeLa

C:Function:

A:Description: may play a role in nuclear protein import

C:Superfamily: nucleoporin Nup358; cyclophilin homology; tetratricopeptide repeat homolo

C:Keywords: leucine zipper

F:26-59/DNA: tetratricopeptide repeat homology &lt;TT1&gt;

F:60-93/DNA: tetratricopeptide repeat homology &lt;TT2&gt;

F:450-471/DNA: leucine zipper #status predicted &lt;LEU&gt;

F:3063-3224/DNA: cyclophilin homology &lt;CTP&gt;

Query Match 41.9%; Score 44; DB 1; Length 3224;

Best Local Similarity 47.4%; Pred. No. 4.4e+02;

Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 2 PAKMSPPGFTVGESEVLS 20

DB 2505 PRAVSPPKFVGSESVKS 2523

RESULT 13

F81444

probable integral membrane protein Cj0263 [imported] - Campylobacter jejuni (strain NCTC

C:Species: Campylobacter jejuni

C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002

C:Accession: F81444

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kettle, J.M.; Churcher, C.; Basham, D.; Chilling

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; Varley, A.; Whitehead, S.; Barrer

Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp

A:Reference number: A81250; PMID:20150912; PMID:10688204

A:Accession: F81444

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-291 &lt;PAR&gt;

A:Cross-references: GB:AL139074; GB:AL111168; NID:96967505; PIDN:CA872721.1; PID:9696774

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: Cj0263

C:Superfamily: gufa protein

Query Match 41.0%; Score 43; DB 2; Length 291;

Best Local Similarity 53.3%; Pred. No. 41;

Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 PAKMSPPGFTVGESE 16

DB 120 POKGQNPFPKFPGEK 134

RESULT 14

AG2399

hypothetical protein all4751 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp. PCC 7120

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 18-Aug-2003

C:Accession: AG2399

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A:Reference number: AB1807; PMID:21595285; PMID:11759640

A:Accession: AG2399

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-305 &lt;KUR&gt;

A:Cross-references: GB:BA000019; PIDN:BA076450.1; PID:gl133888; GSPDB:GNO0179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all4751

C:Superfamily: ATP-MAD kinase

Query Match 41.0%; Score 43; DB 2; Length 305;

Best Local Similarity 30.8%; Pred. No. 44;

Matches 8; Conservative 6; Mismatches 4; Indels 8; Gaps 1;

QY 2 PAKMSPPGFTVGESEVLS 19

DB 56 POKGQNPFPKFPGEK 81

RESULT 15

H70876

hypothetical protein Rv1184c - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C:Accession: H70876

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrer, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; PMID:98295987; PMID:9634230

A:Accession: H70876

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-359 &lt;COL&gt;

A:Cross-references: GB:AL010186; GB:AL123456; NID:93261493; PIDN:CA11861.1; PID:el20395

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: Rv1184c

Query Match 41.0%; Score 43; DB 2; Length 359;

Best Local Similarity 47.8%; Pred. No. 52;

Matches 11; Conservative 3; Mismatches 5; Indels 4; Gaps 2;

QY 1 DPAKMSPPGFTVGESEVLS 19

DB 330 DPAKMSPPGFTVGESEVLS 352

Search completed: March 22, 2004, 07:01:21

Job time : 3.30072 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:31:13 ; Search time 0.767093 Seconds  
(without alignments)  
1357.597 Million cell updates/sec

Title: US-09-662-293-4  
Perfect score: 105  
Sequence: 1 DPAKMSPPRTVEEGVLS 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	ID	Description
1	48	45.7	1 T10B_HUMAN	O14763 homo sapien
2	45	42.9	141 U41_HCV	P16814 human cytom
3	45	42.9	1 ITH3_RAT	O63416 rattus norv
4	44	41.9	1 YORL_TTV	P19296 thermoprote
5	44	41.9	1 BIN4_MOUSE	O92031 mus musculu
6	44	41.9	1 PMA_MYCTU	P96254 mycobacteri
7	44	41.9	1 RBP2_HUMAN	P49792 homo sapien
8	43	41.0	1 ZUP1_CAMEL	O99129 camyllobact
9	43	41.0	1 PPN1_AMASP	O87119 arabidosa sp
10	43	41.0	1 BET3_MESAU	O09029 mesocricetu
11	43	41.0	1 YEL10_HAEIN	P44184 haemophilus
12	43	41.0	1 AAP2_NEUCR	O59392 neurospora
13	43	41.0	1 YG4A_YEAST	P46949 saccharomyc
14	42.5	40.5	1 YNC9_CABEL	P34542 caenorhabdi
15	42.5	40.5	1 CA36_HUMAN	P12111 homo sapien
16	42	40.0	1 STCB_EMEHI	O12608 emericella
17	42	40.0	1 SKD3_MOUSE	O60644 mus musculu
18	42	40.0	1 SKD3_MOUSE	O60644 mus musculu
19	42	40.0	1 L100_ADE40	P11822 human adeno
20	42	40.0	1 ITH3_MESAU	P97280 mesocricetu
21	42	40.0	1 MOP_DESGI	O46509 desulfovibr
22	42	40.0	1 RBP2_BOVIN	P48820 bos taurus
23	41	39.0	1 ENG2_BALSO	O88443 ralsstonia s
24	41	39.0	1 BXB3_CHICK	O90646 gallus gall
25	41	39.0	1 CB22_ORYSA	P12331 oryza sativ
26	41	39.0	1 CB21_ORYSA	P12330 oryza sativ
27	41	39.0	1 LEU2_BUCTU	O55663 synchocyst
28	41	39.0	1 DXR_SYNY3	O55663 synchocyst
29	41	39.0	1 LEU2_BUCTU	O55663 synchocyst
30	41	39.0	1 SR54_ECOLI	O91078 homo sapien
31	41	39.0	1 SR54_ECOLI	O91078 homo sapien
32	41	39.0	1 SR54_ECOLI	O91078 homo sapien
33	41	39.0	1 SR54_ECOLI	O91078 homo sapien

## ALIGNMENTS

RESULT 1  
ID T10B\_HUMAN STANDARD; PRT; 440 AA.  
AC O14763; O14720; O15508; O15517; O15531; Q9BEV0;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Tumor necrosis factor receptor superfamily member 10B precursor (Death receptor 5) (TNF-related apoptosis-inducing ligand receptor 2) (TRAIL receptor 2) (TRAIL-R2).  
GN TNFRSF10B OR DR5 OR TRAILR2 OR TRICK2 OR KILLER OR ZTNFR9.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM LONG), AND SEQUENCE OF N-TERMINUS.  
RC TISSUE=Forebrain fibroblast;  
RX MEDLINE=97459925; PubMed=9311998;  
RA Walczak H., Degli-Esposti M.A., Johnson R.S., Smolak P.J., Faugh J.Y., Bolani N., Timour M.S., Gernhart M.J., Schooley K.A., Smith C.A., Goodwin R.G., Rauch C.T.;  
RA "TRAIL-R2: a novel apoptosis-mediating receptor for TRAIL";  
RL EMBO J. 16:5386-5397(1997).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT), ALTERNATIVE SPLICING, AND VARIANTS LEU-32 AND VAL-67.  
RX MEDLINE=97431692; PubMed=9285725;  
RA Sreaton G.R., Mongkolsapaya J., Xu X.-N., Cowper A.E., McMichael A.J., Bell J.I.;  
RA "TRICK2, a new alternatively spliced receptor that transduces the cytotoxic signal from TRAIL";  
RL Curr. Biol. 7:693-696(1997).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM LONG), CHARACTERIZATION, AND VARIANTS LEU-32 AND VAL-67.  
RX MEDLINE=98039016; PubMed=9373179;  
RA Schneider P., Bodmer J.-L., Thome M., Hofmann K., Holler N., Tschopp J.;  
RA "Characterization of two receptors for TRAIL";  
RL FBS Lett. 416:329-334(1997).  
RN [4]  
RP SEQUENCE FROM N.A. (ISOFORM SHORT).  
RC TISSUE=Ovary;  
RX MEDLINE=97467719; PubMed=9326928;  
RA Wu G.S., Burns T.F., McDonald E.R., III, Jiang W., Meng R., Krantz I.D., Kao G., Gan D.D., Zhou J.Y., Muschel R., Hamilton S.R., Spitzer N.B., Markowitz S., Wu G., el-Deiry W.S.;  
RA "KILLER/DR5 is a DNA damage-inducible p53-regulated death receptor gene";  
RL Nat. Genet. 17:141-143(1997).  
RN [5]  
RP SEQUENCE FROM N.A. (ISOFORM SHORT).  
RX MEDLINE=97390508; PubMed=9242610;  
RA Pan G., Ni J., Wei Y.-F., Yu G.-I., Gentz R., Dixit V.M.;

RT "An antagonist decoy receptor and a death domain-containing receptor  
RT for TRAIL.";  
RT Science 277:815-818(1997).  
[6]  
RP SEQUENCE FROM N.A. (ISOFORM SHORT), AND VARIANTS LEU-32 AND VAL-67.  
RX MEDLINE=97467318; PubMed=9325248;  
RA MacFarlane M., Ahmed M., Srinivasula S.M., Fernandes-Alnemri T.,  
RA Cohen G.M., Alnemri E.S.;  
RT "Identification and molecular cloning of two novel receptors for the  
RT cytotoxic ligand TRAIL.";  
RT J. Biol. Chem. 272:25417-25420(1997).  
[7]  
RP SEQUENCE FROM N.A. (ISOFORM SHORT), AND VARIANTS LEU-32 AND VAL-67.  
RX MEDLINE=98090092; PubMed=9430227;  
RA Chaudhary P.M., Bby M., Jasmin A., Bookwala A., Murray J., Hood L.;  
RT "Death receptor 5, a new member of the TNFR family, and DR4 induce  
RT PAD-dependent apoptosis and activate the NF-kappaB pathway.";  
RT Immunity 7:821-830(1997).  
[8]  
RP SEQUENCE FROM N.A. (ISOFORM SHORT), AND VARIANT LEU-32.  
RX MEDLINE=97390509; PubMed=9242611;  
RA Sheridan J.P., Mersers S.A., Pitti R.M., Gurney A., Skubatch M.,  
RA Baldwin D., Ramakrishnan L., Gray C.L., Baker K., Wood W.I.;  
RT "Control of TRAIL-induced apoptosis by a family of signaling and decoy  
RT receptors.";  
RT Science 277:818-821(1997).  
[9]  
RP SEQUENCE FROM N.A. (ISOFORM LONG), AND VARIANTS LEU-32 AND VAL-67.  
RX Arai T., Akiyama Y., Okabe S., Saito K., Iwai T., Yusa Y.;  
RT "Genomic structure and mutation analyses of the DR5/TRAIL receptor 2  
RT gene in colorectal carcinoma.";  
RT Submitted (May-1998) to the EMBL/Genbank/DBJ databases.  
[10]  
RP SEQUENCE FROM N.A. (ISOFORM SHORT).  
RX Cao X., Zhang W., Wan T.;  
RT Submitted (May-1999) to the EMBL/Genbank/DBJ databases.  
[11]  
RP SEQUENCE FROM N.A. (ISOFORM SHORT), AND VARIANTS LEU-32 AND VAL-67.  
RX Parrish T., Vu T., Gilbert T., Gross J., O'Hara P.;  
RT "Homo sapiens homolog of tumor necrosis factor receptor.";  
RT Submitted (Oct-1999) to the EMBL/Genbank/DBJ databases.  
[12]  
RP SEQUENCE FROM N.A. (ISOFORM LONG), AND VARIANT LEU-32.  
RX TISSUE=Cervix;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang G.M., Hong L.,  
RA Diatchenko L., Manesina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stabileton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raba S.S., Lonnellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richardson S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.M.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A.C., Rodriguez S., Sanchez A.,  
RA Whiting R.M., Madan A.C., Young A.C., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schultz J., Myers R.M.,  
RA Buterfield Y.S.N., Krzywicki M.I., Skalska U., Smalins D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[13]  
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 54-183.  
RX MEDLINE=20017054; PubMed=10549288;  
RA Hymowitz S.G., Christinger H.W., Fuh G., Ulesch M., O'Connell M.,  
RA Kelley R.F., Ashkenazi A., de Vos A.M.;  
RT "Triggering cell death: the crystal structure of Apo2L/TRAIL in a  
RT complex with death receptor 5.";

RT Mol. Cell 4:563-571(1999).  
RN [14]  
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 69-184.  
RX PubMed=10542098;  
RA Mongkolkeha J., Grimes J.M., Chen N., Xu X.-N., Stuart D.I.,  
RA Jones E.Y., Screaton G.R.;  
RT "Structure of the TRAIL-DR5 complex reveals mechanisms conferring  
RT specificity in apoptotic initiation.";  
RT Nat. Struct. Biol. 6:1048-1053(1999).  
[15]  
RP FUNCTION: Receptor for the cytotoxic ligand TNFSF10/TRAIL. The  
CC adaptor molecule FADD recruits caspase-8 to the activated  
CC receptor. The resulting death-inducing signaling complex (DISC)  
CC performs caspase-8 proteolytic activation which initiates the  
CC subsequent cascade of caspases (separate-specific cysteine  
CC proteases) mediating apoptosis. Promotes the activation of NF-  
CC kappaB.  
CC SUBUNIT: Homotrimer. Can interact with TRADD and RIP.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=Long; Synonyms=TRICK2B;  
CC IsoId=O14763-1; Sequence=Displayed;  
CC Name=Short; Synonyms=TRICK2A;  
CC IsoId=O14763-2; Sequence=VSP\_006490;  
CC -1- TISSUE SPECIFICITY: Widely expressed in adult and fetal tissues;  
CC very highly expressed in tumor cell lines such as HeLa S3, K562,  
CC HL-60, SW480, A549 and G361; highly expressed in heart, peripheral  
CC blood lymphocytes, liver, pancreas, spleen, thymus, prostate,  
CC ovary, uterus, placenta, testis, esophagus, stomach and throughout  
CC the intestinal tract; not detectable in brain.  
CC -1- INDUCTION: TNFSF10B is regulated by the tumor suppressor p53.  
CC -1- DISEASE: Defects in TNFSF10B may be a cause of squamous cell  
CC carcinoma of the head and neck (HNSCC) [MIM:601400].  
CC -1- SIMILARITY: Contains 3 TNFR-Cys repeats.  
CC -1- SIMILARITY: Contains 1 death domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AF016849; AAC51778.1; -  
DR EMBL: AF018657; AAB70577.1; -  
DR EMBL: AF018658; AAB70578.1; -  
DR EMBL: AF016266; AAB81180.1; -  
DR EMBL: AF022386; AAB71949.1; -  
DR EMBL: AF012628; AAB67109.1; -  
DR EMBL: AF020501; AAB71412.1; -  
DR EMBL: AF016268; AAC01565.1; -  
DR EMBL: AF012535; AAB67103.1; -  
DR EMBL: AB014715; BAA33723.1; -  
DR EMBL: AB014710; BAA33723.1; JOINED.  
DR EMBL: AB014711; BAA33723.1; JOINED.  
DR EMBL: AB014712; BAA33723.1; JOINED.  
DR EMBL: AB014713; BAA33723.1; JOINED.  
DR EMBL: AB014714; BAA33723.1; JOINED.  
DR EMBL: AB014715; BAA33723.1; JOINED.  
DR EMBL: AB014716; BAA33723.1; JOINED.  
DR EMBL: AB014717; BAA33723.1; JOINED.  
DR EMBL: AF153687; AAF75587.1; -  
DR EMBL: AF192548; AAF07175.1; -  
DR EMBL: BC001281; AAH01281.1; -  
DR PDB: 1D0G; 22-OCT-99.  
DR PDB: 1D4V; 01-NOV-99.  
DR Genew: HGNC:11905; TNFSF10B.  
DR MIM: 603612; -  
DR MIM: 601400; -  
DR GO: GO:0016021; C:integral to membrane; IC.  
DR GO: GO:0016506; F:apoptosis activator activity; NAS.  
DR GO: GO:0008656; F:caspase activator activity; NAS.

Query March 45.7%; Score 48; DB 1; Length 440;  
 Best Local Similarity 43.8%; Pred. No. 7.8;  
 Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 PAKGSPGPIVGEERG 17  
 DB 76 PSEGLCPGHHISDGG 91

## RESULT 2

UL41\_HCMVA STANDARD; PRT; 141 AA.  
 AC P16814;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 01-FEB-1991 (Rel. 17, Last annotation update)  
 DE Hypothetical protein UL41.  
 GN UL41.  
 OS Human cytomegalovirus (strain AD169).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 CC Betaherpesvirinae; Cytomegalovirus.  
 CX NCBI\_TaxID=10360;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90269039; PubMed=2161319;  
 RA Chee M.S., Bankier A.T., Beck S., Bohm R., Brown C.M., Cerny R.,  
 RA Horsnell T., Hutchinson G.A. III, Kouzarides T., Martignetti U.A.,  
 RA Preddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.,  
 RT "Analysis of the protein-coding content of the sequence of human  
 cytomegalovirus strain AD169."  
 RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
 CC EMBL: X17403; CAA35400.1; -  
 DR PIR: S09804; S09804.  
 KW Hypothetical protein.  
 SQ SEQUENCE 141 AA; 1676 MW; 20005377B1EFB712 CRC64;

Query March 42.9%; Score 45; DB 1; Length 141;  
 Best Local Similarity 53.3%; Pred. No. 6.6;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 AKGMSPPGPIVGEERG 17  
 DB 76 AKDLPDGYRGRG 90

## RESULT 3

ITH3\_RAT STANDARD; PRT; 887 AA.  
 AC Q63416;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Inter-alpha-crysein inhibitor heavy chain H3 precursor (ITI heavy  
 DE chain H3) (Inter-alpha-inhibitor heavy chain 3).  
 GN ITH3.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 CX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Liver;  
 RA Blom A., Fries E.,

RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.

CC -1- FUNCTION: MAY ACT AS A CARRIER OF HYALURONAN IN SERUM OR AS A  
 CC BINDING PROTEIN BETWEEN HYALURONAN AND OTHER MATRIX PROTEIN,  
 CC INCLUDING THOSE ON CELL SURFACES IN TISSUES TO REGULATE THE  
 CC LOCALIZATION, SYNTHESIS AND DEGRADATION OF HYALURONAN WHICH ARE  
 CC ESSENTIAL TO CELLS UNDERGOING BIOLOGICAL PROCESSES (BY  
 CC SIMILARITY).  
 CC -1- SUBUNIT: I-alpha-I plasma protease inhibitors are assembled from  
 CC one or two heavy chains (H1, H2 or H3) and one light chain,  
 CC bikunin. Inter-alpha-inhibitor (I-alpha-I) is composed of H1, H2  
 CC and bikunin, inter-alpha-like inhibitor (I-alpha-II) of H2 and  
 CC bikunin, and pre-alpha-inhibitor (P-alpha-I) of H3 and bikunin.  
 CC -1- PTM: Heavy chains are interlinked with bikunin via a chondroitin  
 CC 4-sulfate bridge to the their C-terminal aspartate (By  
 CC similarity).  
 CC -1- SIMILARITY: Belongs to the ITH family.  
 CC -1- SIMILARITY: Contains 1 WFPA domain.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: X83231; CAA58233.1; -  
 DR InterPro: IPR006587; VIT.  
 DR InterPro: IPR002035; VWF\_A.  
 DR Pfam: PF00092; vwa; 1.  
 DR SMART: SM00609; VIT; 1.  
 DR SMART: SM00327; VWA; 1.  
 DR PROSITE: PS50234; WFPA; 1.  
 KW Serine protease inhibitor; Repeat; Signal; Multigene family;  
 KW Glycoprotein.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT PROPEP 22 33 BY SIMILARITY.  
 FT CHAIN 34 647 INTER-ALPHA-TRYPsin INHIBITOR HEAVY CHAIN  
 FT H3.  
 FT PROPEP 648 887 BY SIMILARITY.  
 FT DOMAIN 282 442 WFPA.  
 FT CARBOHYD 91 91 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 580 580 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT BINDING 647 647 CHONDROITIN 4-SULFATE, CROSS-LINK SITE  
 FT (BY SIMILARITY).  
 SQ SEQUENCE 887 AA; 99097 MW; 3B9F0FF96D514096 CRC64;

Query March 42.9%; Score 45; DB 1; Length 887;  
 Best Local Similarity 47.1%; Pred. No. 48;  
 Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 DPAKMSPPGPIVGEERG 17  
 DB 678 DPVGTAVTGQIIIEKG 694

## RESULT 4

YORL\_TTV1 STANDARD; PRT; 232 AA.  
 ID YORL\_TTV1  
 AC P19296;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 01-NOV-1990 (Rel. 16, Last annotation update)  
 DE Hypothetical 26.8 kDa protein.  
 DE Hypothetical 26.8 kDa protein.  
 OS Thermoproteus tenax virus 1 (strain KRA1) (TTV1).  
 OC Viruses; dsDNA viruses, no RNA stage; Lipothirixviridae;  
 OC Lipothirixvirus.  
 CX NCBI\_TaxID=10480;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Neumann H.,  
 RL Submitted (MAR-1989) to the EMBL/Genbank/DBJ databases.



```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

-----

EMBL; X14855; CAA32992.1; -

Hypothetical protein.

SEQUENCE 232 AA; 26862 MW; 0AB2C74AF657CA9D CRC64;

Query Match

Best Local Similarity 54.9%; Score 44; DB 1; Length 232;

Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

8 PGFIVGEGV 18

176 PGGLIGDGI 186

-----

RESULT 5

BIN4\_MOUSE

ID BIN4\_MOUSE STANDARD; PRT; 622 AA.

AC Q9Z0H1;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE WD-repeat protein BIN4.

GN BIN4.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI\_TaxID=10090;

OX [1]

RN [1] SEQUENCE FROM N.A.

RP STRAIN=129/Sv;

RC STRAIN=129/Sv;

RA Rowen L., Qin S., Madan A., Loretz C., Hall J., James R., Doris M.,

RA Shaffer T., Abbasi N., Ratcliffe R., Dickhoff R., Lasky S., Hood L.,

RT "Sequence of the mouse major histocompatibility complex class II

RT region."

RT Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.

RL [2]

RP SEQUENCE FROM N.A.

RC TISSUE=olfactory epithelium;

RX MEDLINE=22386257; PubMed=12477932;

RA Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Datchenko L., Martusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.U., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,

RA Roha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,

RA Bosak S.A., McGwan P.J., McKernan K.J., Malek J.A., Gamaralline P.H.,

RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,

RA Villalón D.K., Muray D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahney J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalios D.E.,

RA Scherich A., Schein J.E., Jones S.U.M., Maitra M.A.,

RT "Generation and initial analysis of more than 15,000 full-length

RT human and mouse cDNA sequences."

RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -1- SIMILARITY: Contains 5 WD repeats.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

-----

EMBL; AF110520; AAC97976.1; -

EMBL; AF100956; AAC69896.1; -

EMBL; BC046977; AAH46977.1; -

MGI; MGI1931871; Bin4.

InterPro: IPR001680; WD40.

Pfam: PF00400; WD40; 1.

SMART: SMO0320; WD40; 3.

PROSITE: PS00678; WD\_REPEATS\_1; FALSE\_NEG.

PROSITE: PS50082; WD\_REPEATS\_2; 1.

PROSITE: PS50294; WD\_REPEATS\_REGION; 1.

Repeat; WD repeat.

REPEAT 192 233 WD 1.

REPEAT 234 271 WD 2.

REPEAT 314 353 WD 3.

REPEAT 356 395 WD 4.

REPEAT 398 435 WD 5.

SEQUENCE 622 AA; 69048 MW; B1BP9A36AC12612 CRC64;

Query Match

Best Local Similarity 41.9%; Score 44; DB 1; Length 622;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

9 PGFIVGEGV 17

156 PGFLVGEDG 164

-----

RESULT 6

PTA\_MYCTU

ID PTA\_MYCTU STANDARD; PRT; 690 AA.

AC P66254;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Phosphate acetyltransferase (EC 2.3.1.8) (Phosphotransacetylase).

GN PTA OR R0408 OR M0421 OR MTCY22G10.04.

OS Mycobacterium tuberculosis.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.

NCBI\_TaxID=1773;

OX [1]

RN [1] SEQUENCE FROM N.A.

RP STRAIN=H37Rv;

RC MEDLINE=98295987; PubMed=9634230;

RX Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,

RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R., Devlin K., Feltham D., Gentles S., Hamlin N., Holtroyd S.,

RA Hornsby T., Jagels K., Krogh A., McLean U., Moule S., Murphy L.,

RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,

RA Sutcliffe J.E., Taylor K., Whitehead S., Barrall B.G.,

RT "Deciphering the biology of Mycobacterium tuberculosis from the

RT complete genome sequence."

RT Nature 393:537-544(1998).

RL [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CDC 1551 / Oshkosh;

RX MEDLINE=22206494; PubMed=12218036;

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,

RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,

RA Kolonyak J.F., Nelson W.C., Umayam L.A., Esmolaeva M., Salzberg S.L.,

RA Delcher A., Utterback T., Weidman J., Khouri H., Kouri J., Mikula A.,

RA Bishel W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;

RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and

RT laboratory strains."

RT J. Bacteriol. 184:5479-5490(2002).

CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + phosphate = CoA + acetyl

CC phosphate.

CC -1- PATHWAY: Conversion of acetate to acetyl-CoA; second step.

```

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; CONTAINS A DOMAIN WITH SOME
CC SIMILARITIES WITH COBO/BIOD.
CC -1- SIMILARITY: In the C-terminal section; belongs to the phosphate
CC acetyltransferase and butyryltransferase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z84724; CAB06578.1; -.
DR EMBL; AE006946; AAK44645.1; -.
DR PIR; F70628; F70628.
DR TIGR; MT0421; -.
DR Tuberculist; Rv0408; -.
DR InterPro; IPR004614; PTA_PTB.
DR InterPro; IPR002505; PTA_PTB.
DR Pfam; PF01515; PTA_PTB; 1.
DR TIGRFAMs; TIGR00651; pta; 1.
DR Transferrase; Acyltransferase; Complete proteome.
KW DOMAIN 365 690 PHOSPHATE ACETYLTRANSFERASE.
FT SEQUENCE 690 AA; 72948 MW; C01C412AF2810CCE CRC64;
SQ
Query Match 41.9%; Score 44; DB 1; Length 690;
Best Local Similarity 40.0%; Pred. No. 52;
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;
QY 1 DPAKMSPPGFIAGEGYLS 20
Db 182 DALRFTPPSYVPEPLTS 201
RESULT 7
RBP2_HUMAN STANDARD; PRT; 3224 AA.
AC P49792; Q15280;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ran-binding protein 2 (RanBP2) (Nuclear pore complex protein Nup358)
DE (Nucleoporin Nup358) (358 kDa nucleoporin) (P270).
GN RANBP2 OR NUP358
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95294031; PubMed=7775481;
RA Wu J., Matunis M.J., Kraemer D., Blobel G., Coutavas E.;
RT "Nup358, a cytoplasmically exposed nucleoporin with peptide repeats,
RT Ran-GTP binding sites, zinc fingers, a cyclophilin A homologous
RT domain, and a leucine-rich region."
RL J. Biol. Chem. 270:14209-14213 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95327194; PubMed=7603572;
RA Yokoyama N., Hayashi N., Seki T., Nishii K., Hayashida T.,
RA Kuma K.I., Miyata T., Fukui M., Nishimoto T., Panté N., Aebi U.;
RT "A giant nucleopore protein that binds Ran/TC4."
RL Nature 376:184-188 (1995).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.96 ANGSTROMS) OF 1171-1304.
RX MEDLINE=99176415; PubMed=10078529;
RA Vetter I.R., Nowak C., Nishimoto T., Kuhlmann J., Wittlinghofer A.;
RT "Structure of a Ran-binding domain complexed with Ran bound to a GTP
RT analogue: implications for nuclear transport."
RL Nature 398:39-46 (1999).
CC -1- FUNCTION: Involved in transport factor (Ran-GTP, karyopherin) -
CC mediated protein import via the P-g repeat-containing domain which
CC acts as a docking site for substrates. Could also have isomerase
CC or chaperone activity and may bind RNA or DNA. Component of the
CC nuclear export pathway. Specific docking site for the nuclear
CC export factor exportin-1.
CC -1- SUBUNIT: Forms a tight complex in association with RANBP1 and the
CC ubiquitin-conjugating enzyme E2 (UBC9) (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear pore complex. Cytoplasmic filaments.
CC -1- DOMAIN: Contains P-X-F-F repeats.
CC -1- SIMILARITY: Contains 8 RanBP2-type zinc fingers.
CC -1- SIMILARITY: Contains 8 RANBP2-type zinc fingers.
CC -1- SIMILARITY: Contains 1 cyclophilin-like ppiase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; LA1840; AAC41758.1; -.
DR EMBL; DA2063; BAA07662.1; -.
DR PIR; S58884; S58884.
DR PDB; IRRP; 1B-MAY-99.
DR Genew; HGNC:9848; RANBP2.
DR MIM; 601181; -.
DR GO; GO:0005643; C:nuclear pore; TAS.
DR GO; GO:0008536; F:RAN protein binding; TAS.
DR GO; GO:0005606; P:protein-nucleus import; TAS.
DR InterPro; IPR002130; CSA_Ppiase.
DR InterPro; IPR000697; EVH1.
DR InterPro; IPR000156; Ran_Bp1.
DR InterPro; IPR008941; TPR-like.
DR InterPro; IPR001440; TPR.
DR InterPro; IPR001876; Znf_RangDP.
DR Pfam; PF00160; pro_isomerase; 1.
DR Pfam; PF00638; Ran_Bp1; 4.
DR Pfam; PF00515; TPR; 1.
DR Pfam; PF00641; ZF-RanBP; 8.
DR PRINTS; PR00153; CSAPPISWASE.
DR SMART; SM00160; RandB; 4.
DR SMART; SMO0547; ZNF RBZ; 8.
DR PROSITE; PS00170; CSA_PPIASE_1; 1.
DR PROSITE; PS00196; RANBD1; 4.
DR PROSITE; PS01358; ZF_RANBP2_1; 8.
DR PROSITE; PS01358; ZF_RANBP2_2; 8.
DR PROSITE; PS01358; ZF_RANBP2_3; 8.
KW Nuclear protein; Transport; Repeat; Zinc-finger; Isomerase; Rotamase;
KW 3D-structure; Polymorphism.
FT DOMAIN 1171 1307
FT ZN_FING 1351 1381 RANBD1 1.
FT ZN_FING 1415 1444 RANBP2-TYPE 1.
FT ZN_FING 1479 1508 RANBP2-TYPE 2.
FT ZN_FING 1543 1572 RANBP2-TYPE 3.
FT ZN_FING 1606 1635 RANBP2-TYPE 4.
FT ZN_FING 1665 1694 RANBP2-TYPE 5.
FT ZN_FING 1724 1753 RANBP2-TYPE 6.
FT ZN_FING 1781 1810 RANBP2-TYPE 7.
FT ZN_FING 1810 1848 RANBP2-TYPE 8.
FT DOMAIN 2012 2148 RANBD1 2.
FT DOMAIN 2309 2445 RANBD1 3.
FT DOMAIN 2911 3046 RANBD1 4.
FT DOMAIN 3067 3223 PPIASE, CYCLOPHILIN-TYPE.
FT VARIANT 1892 1892 P -> A (in dbSNP:12770).
FT VARIANT 1892 1892 /PTG=VAR 014886.
FT CONFLICT 777 777 R -> H (IN REF. 2).
FT CONFLICT 784 784 R -> K (IN REF. 2).
FT STRAND 1191 1204
FT STRAND 1211 1224
FT STRAND 1231 1235
FT TURN 1237 1239
FT STRAND 1242 1244

```

FT STRAND 1247 1247  
 FT STRAND 1255 1255  
 FT TURN 1258 1259  
 FT TURN 1261 1262  
 FT STRAND 1263 1270  
 FT TURN 1272 1273  
 FT STRAND 1277 1284  
 FT HELIX 1288 1300  
 FT TURN 1301 1302  
 SQ SEQUENCE 3224 AA; 358214 MW; 54E78412C96A3C63 CRC64;

Query Match  
 Best Local Similarity 41.9%; Score 44; DB 1; Length 3224;  
 Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 2 PAKGMSPPGFIVGEGVL 20  
 DB 2505 PKAVVSPKVFSGSSVKS 2523

RESULT 8  
 ZUPRT\_CAMJBE STANDARD; PRT; 291 AA.  
 ID ZUPRT\_CAMJBE  
 AC Q8PNT2;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Zinc transporter zuprt.  
 GN ZUPRT OR C10263.  
 OS Campylobacter jejuni.  
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
 OC Campylobacteriaceae; Campylobacter.  
 NC NCB1\_TaxID=157;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCCTC 11168;  
 RX MEDLINE=20150912; PubMed=10688204;  
 RA Parthill U., Wren B.W., Mungall K., Ketley J.M., Churcher C.,  
 RA Basham D., Chillingworth T., Davies R.M., Felwell T., Holroyd S.,  
 RA Jagsels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
 RA Whitehead S., Barrett B.G.;  
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
 RT reveals hypervariable sequences";  
 RL Nature 403:665-668 (2000).  
 CC -!- FUNCTION: Mediates zinc uptake. May also transport other divalent  
 CC cations (by similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (potential).  
 CC -!- SIMILARITY: Belongs to the ZIP transporter (TC 2.A.5) family. Zuprt  
 CC subfamily.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AL139074; CAB72731.1; -.  
 CC PIR; F81444; F81444.  
 DR HAMAP; MF 00548; -!- 1.  
 DR InterPro; IPR003689; Zn\_transp\_1.  
 DR Pfam; PF02335; Zip; 1.  
 KW Transport; Zinc transporter; Transmembrane; Complete proteome.  
 FT TRANSMEM 8 28  
 FT TRANSMEM 39 59  
 FT TRANSMEM 74 94  
 FT TRANSMEM 147 167  
 FT TRANSMEM 174 194  
 FT TRANSMEM 209 229  
 FT TRANSMEM 233 253  
 FT TRANSMEM 271 291  
 FT TRANSMEM POTENTIAL.  
 FT TRANSMEM POTENTIAL.  
 FT TRANSMEM POTENTIAL.

SQ SEQUENCE 291 AA; 31462 MW; 35A0E51B408E1CF2 CRC64;  
 Query Match  
 Best Local Similarity 41.0%; Score 43; DB 1; Length 291;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 2 PAKGMSPPGFIVGEE 16  
 DB 120 POKQNPFFKFPQEK 134

RESULT 9  
 PPN1\_ANASP STANDARD; PRT; 305 AA.  
 ID PPN1\_ANASP  
 AC Q8YNT1;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Probable inorganic polyphosphate/ATP-NAD kinase 1 (EC 2.7.1.23)  
 DE (poly(P)/ATP NAD kinase 1).  
 GN PPNK1 OR ATL4751.  
 OS Arabaena sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 NC NCB1\_TaxID=103690;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=2159285; PubMed=11759840;  
 RA Kaneke T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
 RA Watanabe A., Iritaguchi M., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kohara K., Matsumoto M., Matsuno A., Muraki A.,  
 RA Nishizaki N., Shimpso S., Sugimoto M., Takazawa M., Yamada M.,  
 RA Yasuda M., Tabata S.;  
 RT "Complete genomic sequence of the filamentous nitrogen-fixing  
 RT cyanobacterium Arabaena sp. strain PCC 7120";  
 RL DNA Res. 8:205-213 (2001).  
 CC -!- FUNCTION: Catalyzes the phosphorylation of NAD to NADP. Utilizes  
 CC ATP and other nucleoside triphosphates as well as inorganic  
 CC polyphosphate as a source of phosphorus (by similarity).  
 CC -!- CATALYTIC ACTIVITY: ATP + NAD(+) = ADP + NADP(+).  
 CC -!- COFACTOR: Divalent metal ions (by similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).  
 CC -!- SIMILARITY: Belongs to the NAD kinase family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AP003597; BAB76450.1; -.  
 CC PIR; AG2399; AG2399.  
 DR HAMAP; MF 00361; -!- 1.  
 DR InterPro; IPR002504; ATP\_NADK.  
 DR Pfam; PF01513; NAD\_kinase; 1.  
 KW Transferase; Kinase; NAD; NADP; Complete proteome.  
 SQ SEQUENCE 305 AA; 32798 MW; 028A0D2F3E2A9271 CRC64;

Query Match  
 Best Local Similarity 41.0%; Score 43; DB 1; Length 305;  
 Matches 8; Conservative 6; Mismatches 4; Indels 8; Gaps 1;

Qy 2 PAKGMSPPGF-----IVGEEGVL 19  
 DB 56 PIDGLTPPGFDSWKFAIVLGGDQTV 81

RESULT 10  
 BET3\_MESAU STANDARD; PRT; 367 AA.  
 ID BET3\_MESAU  
 AC O09029;  
 DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE BETA3 protein.  
 OS Mesocricetus auratus (Golden hamster).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 CC Mesocricetus.  
 ON NCBI\_TaxID=10036;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96140430; PubMed=8552091;  
 RA Peyton M., Sellrecht C.M.M., Naya F.J., Huang H.-P., Samora P.J.,  
 RA Tsai M.-U.;  
 RT "BETA3, a novel helix-loop-helix protein, can act as a negative  
 RT regulator of BETA2 and MyoD-responsive genes";  
 RL Mol. Cell. Biol. 16:626-633(1996).  
 CC -1- FUNCTION: Inhibits DNA binding of TCF3 (E47) homodimers and TCF3  
 CC (E47)/NEUROD1 heterodimers and acts as a strong repressor of  
 CC Neurod1 and MyoD-responsive genes, probably by heterodimerization  
 CC with class a basic helix-loop-helix factors. Despite the presence  
 CC of an intact basic domain, does not bind to DNA.  
 CC -1- SUBUNIT: HETERODIMER WITH OTHER BHLH PROTEINS, LIKE TCF3 (E47).  
 CC -1- SUBCELLULAR LOCATION: Nuclear, lung, brain and pancreas (insulinoma).  
 CC -1- TISSUE SPECIFICITY: Kidney, lung, brain and pancreas (insulinoma).  
 CC -1- SIMILARITY: Contains 1 basic helix-loop-helix (BHLH) domain.  
 CC -----  
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; S80870; AAB50691.1; -.  
 CC DR TRANSPAC; T01674; -.  
 CC DR InterPro; IPR001092; HLH\_basic.  
 CC DR Pfam; PF00010; HLH; 1.  
 CC DR SMART; SM00353; HLH; 1.  
 CC DR PROSITE; PS50888; HLH; 1.  
 CC KW Nuclear protein; Transcription regulation; Repressor.  
 CC FT DOMAIN 11 14 POLY-ALA.  
 CC FT DOMAIN 58 62 POLY-SER.  
 CC FT DOMAIN 83 99 POLY-GLY.  
 CC FT DOMAIN 174 179 POLY-GLY.  
 CC FT DOMAIN 204 217 POLY-GLY.  
 CC FT DNA BIND 229 240 BASIC DOMAIN.  
 CC FT DOMAIN 241 282 HELIX-LOOP-HELIX MOTIF.  
 CC FT DOMAIN 311 319 POLY-ALA.  
 CC SQ SEQUENCE 367 AA; 35905 MW; 6CAB9AF96B85F77 CRC64;  
 Query Match 41.0%; Score 43; DB 1; Length 367;  
 Best Local Similarity 57.1%; Pred. No. 38;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 QY 5 GMSPPGPIVGEVY 18  
 Db 99 GVSVPGLVGSAGV 112  
 RESULT 11  
 YELI\_HAEIN STANDARD; PRT; 394 AA.  
 AC P44184;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Protein H11410.  
 GN H11410.  
 OS Haemophilus influenzae.  
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 CC Pasteurellaceae; Haemophilus.  
 ON NCBI\_TaxID=727;  
 CC

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=Rd / KM20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Sadek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,  
 RA Gresh C.L., McDonald J.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 RT Rd";  
 RL Science 269:496-512(1995).  
 RN [2]  
 RP IDENTIFICATION BY MASS SPECTROMETRY.  
 RX MEDLINE=20137488; PubMed=10675023;  
 RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,  
 RA Gray C., Fountoulakis M.;  
 RT "Two-dimensional map of the proteome of Haemophilus influenzae";  
 RL Electrophoresis 21:411-429(2000).  
 CC -----  
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; U32820; AAC23058.1; -.  
 CC DR PIR; E64028; E64028.  
 CC DR TIGR; H11410; -.  
 CC DR InterPro; IPR006701; G2P.  
 CC DR InterPro; IPR006437; Phage\_term\_2.  
 CC DR Pfam; PF04466; G2P; 1.  
 CC DR TIGRFAMs; TIGR01547; phage\_term\_2; 1.  
 CC KW Complete proteome.  
 CC SQ SEQUENCE 394 AA; 44782 MW; D4217986B8A6C772 CRC64;  
 Query Match 41.0%; Score 43; DB 1; Length 394;  
 Best Local Similarity 60.0%; Pred. No. 41;  
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
 QY 3 AKGMSPPGPIVGEV 17  
 Db 186 AKGMSPPGPIVGEV 200  
 RESULT 12  
 AAP2\_NEUCR STANDARD; PRT; 551 AA.  
 ID AAP2\_NEUCR  
 AC 059942;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Amino acid permease 2.  
 GN AAP-2 OR AAP2.  
 OS Neurospora crassa.  
 CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 CC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.  
 ON NCBI\_TaxID=5141;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Margolles-Clark E., Bowman B.U.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (probable).  
 CC -1- SIMILARITY: Belongs to the amino acid permease family.  
 CC -----  
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; AF053231; AAC08355.1; -  
DR InterPro; IPR002293; AA/rel\_permease.  
DR InterPro; IPR004756; AA/permease.  
DR InterPro; IPR004840; AAC/permease.  
DR InterPro; IPR004841; Permease\_region.  
DR Pfam; PF00324; aa\_permeases; 1.  
DR TIGRfam; TIGR00907; 2A0304; 1.  
DR PROSITE; PS00218; AMINO ACID PERMEASE 1; 1.  
KM Transport; Amino-acid transport; Transmembrane.  
FT TRANSMEM 66 86 POTENTIAL.  
FT TRANSMEM 90 110 POTENTIAL.  
FT TRANSMEM 139 159 POTENTIAL.  
FT TRANSMEM 188 208 POTENTIAL.  
FT TRANSMEM 214 234 POTENTIAL.  
FT TRANSMEM 255 275 POTENTIAL.  
FT TRANSMEM 301 321 POTENTIAL.  
FT TRANSMEM 347 367 POTENTIAL.  
FT TRANSMEM 399 419 POTENTIAL.  
FT TRANSMEM 424 444 POTENTIAL.  
FT TRANSMEM 464 484 POTENTIAL.  
FT TRANSMEM 496 516 POTENTIAL.  
SQ SEQUENCE 551 AA; 59869 MW; C812C646882F9ADF CRC64;

Query Match  
Best Local Similarity 41.0%; Score 43; DB 1; Length 551;  
Matches 7; Conservativity 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 PAKGSPGPGF 11  
DB 252 PSSGWSPPGF 261

RESULT 13  
Y4A\_YEAST STANDARD; PRT; 817 AA.  
ID P46949;  
AC 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical 90.8 kDa protein in RRP4-SNG1 intergenic region.  
GN YGR196C OR G7589.

OS Saccharomyces cerevisiae (Baker's Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C;  
RX MEDLINE=96076633; PubMed=7502584;  
RA Guerrero I., Mala e Silva A., Barreiros T., Arroyo J.,  
RA Garcia-Conalaz M., Garcia-Saez M.I., Rodriguez-Pousada C.,  
RA Nombela C.;  
RT "The complete sequence of a 9000 bp fragment of the right arm of  
RT Saccharomyces cerevisiae chromosome VII contains four previously  
RT unknown open reading frames."  
RL Yeast 11:1087-1091 (1995).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X82775; CAA58019.1; -  
DR EMBL; Z72961; CAA97222.1; -

DR PIR; S53919; S53919.  
DR GeneMol; 141508; -  
DR SCD; S0003428; FYV8.  
DR GO; GO:0005737; Cytoplasm; IDA.  
KM Hypothetical protein.  
SQ SEQUENCE 817 AA; 90797 MW; E52C5D659D63BEBB CRC64;

Query Match  
Best Local Similarity 41.0%; Score 43; DB 1; Length 817;  
Matches 7; Conservativity 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 DPAKSPGPGTVEGCVL 19  
DB 547 DTANKKAPPGVYDNGKL 565

RESULT 14  
YNC9\_CABEL STANDARD; PRT; 646 AA.  
ID YNC9\_CABEL  
AC P34542; P34543;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical protein R05D3.9 in chromosome III precursor.  
GN R05D3.9/R05D3.10.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,  
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,  
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rinken L., Roopra A., Saunders D., Showkeen R.,  
RA Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,  
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
RA Wohldman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans."  
RL Nature 368:32-38 (1994).

RN [2]  
RP REVISIONS.  
RA Waterston R.;  
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; L07144; AKK21444.2; -  
DR PIR; S44863; S44863.  
DR PIR; S44865; S44865.  
DR WormPep; R05D3.9; CE31049.  
KM Hypothetical protein; Signal.  
FT SIGNAL 1 29 POTENTIAL.  
FT CHAIN 30 646 HYPOTHETICAL PROTEIN R05D3.9.  
SQ SEQUENCE 646 AA; 72631 MW; 887932547DB5B141 CRC64;

Query Match  
Best Local Similarity 40.5%; Score 42.5; DB 1; Length 646;  
Matches 10; Conservativity 1; Mismatches 5; Indels 1; Gaps 1;

QY 2 PAKGMSPPG-FIVESEG 17  
 DB 42 PPSGSPSPKFPDDEG 58

RESULT 15  
 CA36 HUMAN  
 ID CA36 HUMAN STANDARD; PRT; 3176 AA.  
 AC P12111; 016501;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE COL6A3  
 GN COL6A3  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 CX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Fibroblast;  
 RX MEDLINE=90151612; PubMed=1689238;  
 RA Chu M.-L., Zhang R.-Z., Pan T.-C., Stokes D., Conway D., Kuo H.-J.,  
 RA Glanville R., Meyer U., Mann K., Deutzmann R., Timpl R.;  
 RT "Mosaic structure of globular domains in the human type VI collagen  
 alpha 3 chain: similarity to von Willebrand factor, fibronectin,  
 RT actin, salivary proteins and apocytin type protease inhibitors.";  
 RL EMBO J. 9:385-393(1990).  
 RN [2]  
 RP REVISIONS.  
 RA Chu M.-L.;  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 2038-2373 FROM N.A.  
 RX MEDLINE=89066644; PubMed=3198591;  
 RA Chu M.-L., Conway D., Pan T.-C., Baldwin C., Mann K., Deutzmann R.,  
 RA Timpl R.;  
 RT "Amino acid sequence of the triple-helical domain of human collagen  
 type VI.";  
 RL J. Biol. Chem. 263:18601-18606(1988).  
 RN [4]  
 RP SEQUENCE OF 2092-2157 FROM N.A.  
 RX MEDLINE=88029444; PubMed=365927;  
 RA Chu M.-L., Mann K., Deutzmann R., Pribula-Conway D.,  
 RA Hsu-Chen C.-C., Bernard M.P., Timpl R.;  
 RT "Characterization of three constituent chains of collagen type VI by  
 RT peptide sequences and cDNA clones.";  
 RL Eur. J. Biochem. 168:309-317(1987).  
 RN [5]  
 RP SEQUENCE OF 2092-2151 FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=88161046; PubMed=3348212;  
 RA Weil D., Matzel M.-G., Passage E., van Cong N., Pribula-Conway D.,  
 RA Mann K., Deutzmann R., Timpl R., Chu M.-L.;  
 RT "Cloning and chromosomal localization of human genes encoding the  
 RT three chains of type VI collagen.";  
 RL Am. J. Hum. Genet. 42:435-445(1988).  
 RN [6]  
 RP SEQUENCE OF 32-236 FROM N.A., AND ALTERNATIVE SPLICING.  
 RX MEDLINE=93054780; PubMed=1339440;  
 RA Zanusai S., Doliana R., Segat D., Bonaldo P., Colombatti A.;  
 RT "The human type VI collagen gene, mRNA and protein variants of the  
 RT alpha 3 chain generated by alternative splicing of an additional 5-end  
 RT exon.";  
 RL J. Biol. Chem. 267:24082-24089(1992).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 3107-3164.  
 RX MEDLINE=95182468; PubMed=7533217;  
 RA Airoux B., Merieau K., Saludjian P., Norris F., Norris K., Bioern S.,  
 RA Olsen O., Petersen L., Ducruix A.;  
 RT "The 1.6 A structure of Kunitz-type domain from the alpha 3 chain of  
 RT human type VI collagen.";  
 RL J. Mol. Biol. 246:609-617(1995).

RN [8]  
 RP STRUCTURE BY NMR OF 3102-3164.  
 RX MEDLINE=96398604; PubMed=8805527;  
 RA Zweckstetter M., Czisch M., Mayer U., Chu M.-L., Zinth W., Timpl R.,  
 RA Holak T.A.;  
 RT "Structure and multiple conformations of the Kunitz-type domain from  
 RT human type VI collagen alpha3 (VI) chain in solution.";  
 RL Structure 4:195-209(1996).  
 RN [9]  
 RP STRUCTURE BY NMR OF 3107-3164.  
 RX MEDLINE=97410311; PubMed=9265624;  
 RA Soerensen M.D., Bioern S., Norris K., Olsen O., Petersen L.,  
 RA James T.L., Led U.J.;  
 RT "Solution structure and backbone dynamics of the human alpha3-chain  
 RT type VI collagen C-terminal Kunitz domain.";  
 RL Biochemistry 36:10439-10450(1997).  
 RN [10]  
 RP DISEASE.  
 RX MEDLINE=21987636; PubMed=11992252;  
 RA Demir E., Sabatelli P., Allmand V., Ferreira A., Moghadaszadeh B.,  
 RA Makrelouf M., Topaloglu H., Echeine B., Merini L., Gutcheney P.;  
 RT "Mutations in COL6A3 cause severe and mild phenotypes of Ullrich  
 RT congenital muscular dystrophy.";  
 RL Am. J. Hum. Genet. 70:1446-1458(2002).  
 RN [11]  
 RP VARIANT BM GLU-1679, AND VARIANT HIS-2831.  
 RX MEDLINE=98204804; PubMed=9536084;  
 RA Pan T.-C., Zhang R.-Z., Pericak-Vance M.A., Tandan R., Fries T.,  
 RA Stajich J.M., Viles K., Vance J.M., Chu M.-L., Speer M.C.;  
 RT "Missense mutation in a von Willebrand factor type A domain of the  
 RT alpha 3(VI) collagen gene (COL6A3) in a family with Bethlem  
 RT myopathy.";  
 RL Hum. Mol. Genet. 7:807-812(1998).  
 CC -1- FUNCTION: Collagen VI acts as a cell-binding protein.  
 CC -1- SUBUNIT: Trimers composed of three different chains: alpha 1(VI),  
 CC alpha 2(VI), and alpha 3(VI).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC Name=2;  
 CC IsoId=P12111-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=P12111-2; Sequence=VSP\_001172;  
 CC -1- PTM: Prolines at the third position of the tripeptide repeating  
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
 CC -1- DISAAS: Defects in COL6A3 are a cause of Bethlem myopathy (BM)  
 CC (MIM:158810). BM is a rare autosomal dominant proximal myopathy  
 CC characterized by early childhood onset (complete penetrance by the  
 CC age of 5) and joint contractures most frequently affecting the  
 CC elbows and ankles.  
 CC -1- DISEASE: Defects in COL6A3 are a cause of Ullrich congenital  
 CC muscular dystrophy (UCMD) (MIM:254090); also known as Ullrich  
 CC scleroatonic muscular dystrophy. UCMD is an autosomal recessive  
 CC congenital myopathy characterized by muscle weakness and multiple  
 CC joint contractures, generally noted at birth or early infancy. The  
 CC clinical course is more severe than in Bethlem myopathy.  
 CC -1- SIMILARITY: Contains 1 BPT/Kunitz inhibitor domain.  
 CC -1- SIMILARITY: Contains 1 fibronectin type III domain.  
 CC -1- SIMILARITY: Contains 12 WPA domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X52022; CA36267.1; -;  
 DR EMBL: X06196; CA36267.1; -;  
 DR EMBL: M20778; -; NOT ANNOTATED\_CDS.  
 DR EMBL: M27449; AA52057.1; -;  
 DR EMBL: S49432; AA324261.1; -;  
 DR FIR: AS9140; CGH03A.

DR PDB: 1KNT; 01-NOV-94.  
 DR PDB: 2KNT; 15-MAY-97.  
 DR PDB: 1KUN; 12-NOV-97.  
 DR PDB: 1KTH; 28-AUG-02.  
 DR Genew; HGNC:2213; COL6A3.  
 DR MIM: 120250; -.  
 DR MIM: 158810; -.  
 DR MIM: 254090; -.  
 DR GO; GO:0005589; C:collagen type VI; TAS.  
 DR GO; GO:0007517; P:muscle development; TAS.  
 DR InterPro; IPR008161; Clg helix.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR003861; FN\_III.  
 DR InterPro; IPR002223; Kunitz\_BPTI.  
 DR Pfam; PF01391; Collagen; 5.  
 DR Pfam; PF00014; Kunitz\_BPTI; 1.  
 DR Pfam; PF00092; vwa; 1.  
 DR PRINTS; PR00759; BASICPTASE.  
 DR PRINTS; PR00453; VMPADOMAIN.  
 DR ProDom; PD000007; Clg helix; 2.  
 DR ProDom; PD000222; Kunitz\_BPTI; 1.  
 DR SMART; SM00060; FN3; 1.  
 DR SMART; SM00131; KU; 1.  
 DR SMART; SM00327; VWA; 12.  
 DR PROSITE; PS00280; BPTI\_KUNITZ\_1; 1.  
 DR PROSITE; PS00279; BPTI\_KUNITZ\_2; 1.  
 DR PROSITE; PS50234; VMPA; 12.  
 KM Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KM Glycoprotein; Cell adhesion; Serine protease inhibitor; Collagen;  
 KM Signal; 3D-structure; Disease mutation; Polymorphism;  
 KM Alternative splicing.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 3176 COLLAGEN ALPHA 3(VI) CHAIN.  
 FT DOMAIN 26 2038 NONHELICAL REGION.  
 FT DOMAIN 2039 2375 TRIPLE-HELICAL REGION.  
 FT DOMAIN 2376 3176 NONHELICAL REGION.  
 FT DOMAIN 39 213 VMPA 1.  
 FT DOMAIN 242 419 VMPA 2.  
 FT DOMAIN 445 620 VMPA 3.  
 FT DOMAIN 639 816 VMPA 4.  
 FT DOMAIN 837 1009 VMPA 5.  
 FT DOMAIN 1029 1205 VMPA 6.  
 FT DOMAIN 1233 1404 VMPA 7.  
 FT DOMAIN 1436 1609 VMPA 8.  
 FT DOMAIN 1639 1812 VMPA 9.  
 FT DOMAIN 1838 2024 VMPA 10.  
 FT DOMAIN 2402 2581 VMPA 11.  
 FT DOMAIN 2619 2815 VMPA 12.  
 FT DOMAIN 2987 3076 FIBRONECTIN TYPE-III.  
 FT DOMAIN 3107 3176 BPTI/KUNITZ INHIBITOR.  
 FT SITE 2040 2042 CELL ATTACHMENT SITE.  
 FT SITE 2136 2138 CELL ATTACHMENT SITE.  
 FT SITE 2148 2150 CELL ATTACHMENT SITE.  
 FT SITE 2154 2156 CELL ATTACHMENT SITE.  
 FT SITE 2370 2372 CELL ATTACHMENT SITE.  
 FT ACT SITE 3121 3122 REACTIVE BOND.  
 FT DISULFID 3120 3144  
 FT DISULFID 3136 3157  
 FT CARBOHYD 202

N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 40.5%; Score 42.5; DB 1; Length 3176;  
 Best Local Similarity 47.1%; Pred. No. 4.5e+02;  
 Matches 8; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

QY 2 PAKGMSPPGFIYGESEV 18  
 DB 2232 PAKGAPPG-LIGEGGI 2247

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:39:53 / Search time 4.0246 Seconds  
(without alignments)  
1568.003 Million cell updates/sec

Title: US-09-662-293-4  
Perfect score: 105  
Sequence: 1 DPAXGMSPPGFVIGSEGVLS 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_proteint:\*  
12: sp\_virus:\*  
13: sp Vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rv1rus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	99	94.3	555	5	Q9UER7 dermatophag
2	50	47.6	408	5	Q8MRZ9
3	50	47.6	408	5	Q8MRZ9 drosophila
4	48	45.7	118	4	Q72360
5	47	44.8	201	16	Q82S72
6	47	44.8	573	10	Q8X141
7	47	44.8	584	10	Q8RXF2
8	47	44.8	585	10	Q8H1F3
9	47	44.8	880	16	Q8H7U2
10	47	44.8	880	16	Q8B2D5
11	46.5	44.3	198	2	Q7X4K8
12	46	43.8	144	2	Q8RTY9
13	46	43.8	433	16	Q9A5K3
14	46	43.8	455	10	Q8H849
15	46	43.8	662	17	Q974C3
16	46	43.8	3075	13	Q8AW10

17	46	43.8	4641	4	Q75592	Q75592 homo sapien
18	46	43.8	4708	11	Q7TPH6	Q7TPH6 mus musculu
19	45	42.9	79	16	Q8BUT1	Q8BUT1 rhizobium 1
20	45	42.9	115	12	Q39919	Q39919 human cytom
21	45	42.9	174	16	Q8A3Y9	Q8A3Y9 bacteroides
22	45	42.9	283	16	Q8A1P1	Q8A1P1 bacteroides
23	45	42.9	303	16	Q8NMD1	Q8NMD1 corynebacte
24	45	42.9	351	16	Q8PEL3	Q8PEL3 xanthomonas
25	45	42.9	353	5	Q8MZL0	Q8MZL0 drosophila
26	45	42.9	402	11	Q91Z98	Q91Z98 mus musculu
27	45	42.9	402	11	Q8VH43	Q8VH43 mus musculu
28	45	42.9	424	16	Q89RH4	Q89RH4 bradyrhizob
29	45	42.9	494	2	Q92N72	Q92N72 aeromonas s
30	45	42.9	625	16	Q8D977	Q8D977 vibrio vuln
31	45	42.9	634	16	Q87P74	Q87P74 vibrio para
32	45	42.9	635	16	Q92KX0	Q92KX0 rhizobium m
33	45	42.9	785	5	Q7YDE6	Q7YDE6 cryptocospori
34	44.5	42.4	6298	11	Q8VH43	Q8VH43 mus musculu
35	44	41.9	111	17	Q9VAE9	Q9VAE9 mus musculu
36	44	41.9	159	17	Q8PXB0	Q8PXB0 methanosa
37	44	41.9	256	5	Q9NCB5	Q9NCB5 tritibolium c
38	44	41.9	306	16	Q87M92	Q87M92 vibrio para
39	44	41.9	342	4	Q13074	Q13074 homo sapien
40	44	41.9	375	3	Q873B0	Q873B0 neurospora
41	44	41.9	403	17	Q8ZXI6	Q8ZXI6 pyrobaculum
42	44	41.9	405	17	Q87ZU1	Q87ZU1 pyrococcus
43	44	41.9	408	17	Q97ZM0	Q97ZM0 sulfobolus
44	44	41.9	410	17	Q9VAE1	Q9VAE1 aeropyrum p
45	44	41.9	444	16	Q981I7	Q981I7 rhizobium 1

#### ALIGNMENTS

RESULT 1  
Q9UER7 PRELIMINARY; PRT; 555 AA.  
AC Q9UER7  
DT 01-MAY-2000 (TRENDELREL. 13, Created)  
DT 01-MAY-2000 (TRENDELREL. 13, Last sequence update)  
DT 01-OCT-2003 (TRENDELREL. 25, Last annotation update)  
DE 98kDa HDV allergen.  
OS Dermatophagoides farinae (House-dust mite).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
OC Acariformes; Sarcophormes; Astigmata; Psoroptida; Analgoidea;  
OC Pyroglyphidae; Dermatophagoides.  
OX NCBI\_TaxID=6594;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Weber B.R., Hunter S., Steadman K., McCall C.;  
RT "Cloning and Characterization of a 98 kDa Allergen from  
RT Dermatophagoides farinae.", EMBL/GenBank/DBJ databases.  
RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF178772; AAD52672.1; -  
DR GO; GO:0005576; C:extracellular; IEA.  
DR GO; GO:0008061; F:chitin binding; IEA.  
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR GO; GO:0006030; P:chitin metabolism; IEA.  
DR InterPro; IPR002557; Chitin bind. Per.  
DR InterPro; IPR001223; Glyco\_hydro\_18.  
DR InterPro; IPR001579; Glyco\_hydro\_18A.  
DR Pfam; PF00704; Glyco\_hydro\_18; 1.  
DR ProDom; PD000471; Glyco\_hydro\_18; 1.  
DR SMART; SM00494; ChitBD2; 1.  
DR SMART; SM00636; Glyco\_18; 1.  
DR PROSITE; PS01095; CHITINASE\_18; 1.  
KW Glycosidase; Hydrolase.  
SQ SEQUENCE 555 AA; 63238 MW; 0B45641A45B30B CRC64;  
Query Match 94.3%; Score 99; DB 5; Length 555;  
Best Local Similarity 95.0%; Pred. No. 3.7e-07;  
Matches 19; Conservative 0; Mismatches 1; Indels 0;



QY 1 DPAKMSPPGFIVGEGVLS 20  
 Db 298 DPAKMSPPGFIVGEGVLS 317

## RESULT 2

Q8MRZ9 PRELIMINARY; PRT; 408 AA.  
 AC Q8MRZ9;  
 DT 01-OCT-2002 (TRENBLrel. 22, Created)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE RH61522P.  
 GN EG:115C2.8 OR CG13367.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_Taxid=7227;  
 RN NCBI\_Taxid=7227;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkely;  
 RA Stapleton M., Brackstein P., Hong L., Agbayan A., Carlson J.,  
 RA Chavez M., Chavez C., Dorsett V., Dresnek D., Farran D., Frise E.,  
 RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,  
 RA Miranda A., Mungall C.J., Nuno J., Pacle J., Paragas V., Park S.,  
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
 RA Celniker S.;  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY119176; AAM51036.1; -;  
 DR FLYbase; FBgn0025634; EG:115C2.8.  
 SQ SEQUENCE 408 AA; 43953 MW; 0FD51AE3C6303960 CRC64;

Query Match 47.6%; Score 50; DB 5; Length 408;  
 Best Local Similarity 40.0%; Pred. No. 16;  
 Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;  
 QY 1 DPAKMSPPGFIVGEGVLS 20  
 Db 327 DPAKMSPPGFIVGEGVLS 346

## RESULT 3

Q9WSE2 PRELIMINARY; PRT; 408 AA.  
 AC Q9WSE2; 077428;  
 DT 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE EG:115C2.8 protein.  
 GN EG:115C2.8 OR CG13367.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_Taxid=7227;  
 RN NCBI\_Taxid=7227;  
 RP SEQUENCE FROM N.A.  
 RA Sallies C., Valenti P., Darlamitsou A., Henderson N., Campbell L.,  
 RA Glover D.;  
 RL "Sequencing the distal X chromosome of Drosophila melanogaster";  
 RT Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Benos P.;  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE003418; AAF4535.1; -;  
 DR EMBL; AL031581; CAA20887.1; -;  
 DR PIR; T13387; T13387.  
 DR FLYbase; FBgn0025634; EG:115C2.8.  
 SQ SEQUENCE 408 AA; 43939 MW; 8B9E0E52E086F4F CRC64;

Query Match 47.6%; Score 50; DB 5; Length 408;  
 Best Local Similarity 40.0%; Pred. No. 16;  
 Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 DPAKMSPPGFIVGEGVLS 20  
 Db 327 DPAKMSPPGFIVGEGVLS 346

## RESULT 4

Q7Z360 PRELIMINARY; PRT; 118 AA.  
 AC Q7Z360;  
 DT 01-OCT-2003 (TRENBLrel. 25, Created)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Hypothetical protein DKFZp686I01145.  
 GN DKFZp686I01145.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Human colon endothel primary cell culture;  
 RA Bloecher H., Boecher M., Mewes H.W., Weil B., Amd C., Osanger A.,  
 RA Fobo G., Han M., Wiemann S.;  
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BX538104; CAD98017.1; -;  
 KM Hypothetical protein.  
 SQ SEQUENCE 118 AA; 12808 MW; 0310EF37E1F0E92C CRC64;

Query Match 45.7%; Score 48; DB 4; Length 118;  
 Best Local Similarity 43.8%; Pred. No. 9;  
 Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 PAKMSPPGFIVGEGVLS 17  
 Db 76 PSEGLCPGHHSDEG 91

## RESULT 5

Q82S22 PRELIMINARY; PRT; 201 AA.  
 AC Q82S22;  
 DT 01-JUN-2003 (TRENBLrel. 24, Created)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN NE2330.  
 OS Nitrosomonas europaea.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;  
 OC Nitrosomonadaceae; Nitrosomonas.  
 NCBI\_Taxid=915;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19718 / IFO 14228;  
 RX MEDLINE=22586410; PubMed=12700255;  
 RA Chain P., Lamerdin J.B., Larimer F.W., Regala W., Lao V., Land M.,  
 RA Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,  
 RA Arciero D.M., Holmes N.G., Whitaker M.M., Ard D.J.;  
 RT "Complete genome sequence of the ammonia-oxidizing bacterium and  
 RT obligate chemolithoautotroph Nitrosomonas europaea";  
 RL J. Bacteriol. 185:2759-2773(2003).  
 DR EMBL; BX21864; CAD86242.1; -;  
 DR InterPro; IPR005572; RecA\_N.  
 DR Pfam; PF03872; RecA\_N.  
 KM Hypothetical protein, Complete proteome.  
 SQ SEQUENCE 201 AA; 22079 MW; C5364D2B5BCD1A45 CRC64;

Query Match 44.8%; Score 47; DB 16; Length 201;  
 Best Local Similarity 47.4%; Pred. No. 23;  
 Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;





```

DE Aspartate transcarbamoylase, pyrC' subunit.
GN CC2444.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Land M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Pocko A.I., Nelson W.C., Newton A., Stephens C., Shadke N.D., Ely B.,
RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Ureiderack T., Tan K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL: AE005913; AX24415.1;
DR PIR: C87552; C87552.
DR TIGR: CC2444;
DR GO: GO:0016787; F:hydrolase activity; IEA.
DR InterPro: IPR006680; Amidohydro 1.
DR InterPro: IPR004722; Pept_M38_DHOMult.
DR InterPro: IPR005847; Pept_M38_Regn.
DR Pfam: PF01979; Amidohydro_1; 1.
DR PRODOM: PD000518; Urease; 1.
DR TIGRFAMs: TIGR00857; pyrC_multi; 1.
KM Complete proteome.
SQ SEQUENCE 433 AA; 45438 MW; 85E4B02BB00A9A1A CRC64;

Query Match 43.8%; Score 46; DB 16; Length 433;
Best Local Similarity 40.0%; Pred. No. 76;
Matches 8; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 DPAKMSPPGFIVGEEVLT 20
DB 16 DPEBGYDGPVGIVSEGVIT 35

RESULT 14
QY 08H849 PRELIMINARY; PRT; 455 AA.
AC 08H849;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative glycosyl hydrolase.
GN OJ1626B05.8.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Euphorbiaceae; Oryzae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Wing R.A., Yu Y., Soderlund C., Kim H.-R., Rambo T., Saeki C.,
RA Currie J., Collier K.;
RT "Rice Genomic Sequence.";
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC104473; AAN60993.1;
DR GO: GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO: GO:0016787; F:hydrolase activity; IEA.
DR GO: GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro: IPR000490; Glyco_hydro_17.
DR Pfam: PF00333; Glyco_hydro_17; 1.
KM Hydrolase.
SQ SEQUENCE 455 AA; 47499 MW; 50B83788BED132BB CRC64;

Query Match 43.8%; Score 46; DB 10; Length 455;
Best Local Similarity 47.1%; Pred. No. 80;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 2 DPAKMSPPGFIVGEEV 18
DB 391 PSRGMPTGCVTAAGV 407

RESULT 15
QY 0974C3 PRELIMINARY; PRT; 662 AA.
AC 0974C3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative acetyl-CoA synthetase.
GN ST0730.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / ?;
RX MEDLINE=21456156; PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Anka A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
DR EMBL: AP000983; BAB5737.1;
DR GO: GO:0003824; F:catalytic activity; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro: IPR000873; AMP-bind.
DR Pfam: PF00501; AMP-binding; 1.
DR PROSITE: PS00455; AMP-BINDING.
DR PROSITE: PS00455; AMP-BINDING.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 662 AA; 74544 MW; FC698FB9A2C107 CRC64;

Query Match 43.8%; Score 46; DB 17; Length 662;
Best Local Similarity 44.4%; Pred. No. 1,2e+02;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 DPAKMSPPGFIVGEEV 18
DB 561 DPKGEVFAVFLKGV 578

Search completed: March 22, 2004, 06:59:13
Job time : 6.0246 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:30:23 ; Search time 6.0816 Seconds  
(Without alignments)  
929.256 Million cell updates/sec

Title: US-09-662-293-4  
Perfect score: 105  
Sequence: 1 DPAKMSPPGFIVGESVLS 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29fnd04:.\*  
1: geneseqp1980s:.\*  
2: geneseqp1990s:.\*  
3: geneseqp2000s:.\*  
4: geneseqp2001s:.\*  
5: geneseqp2002s:.\*  
6: geneseqp2003as:.\*  
7: geneseqp2003bs:.\*  
8: geneseqp2004s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	100.0	20	AAV52513	Aay52513 House dus
2	105	100.0	20	AAV52522	Aay52522 House dus
3	105	100.0	20	AAU96317	Aau96317 Der HMW-m
4	105	100.0	20	AAU96326	Aau96326 Der HMW-m
5	101	96.2	40	AAV52535	Aay52535 D. pteron
6	101	96.2	49	AAU96339	Aau96339 Der HMW-m
7	101	96.2	50	AAV52533	Aay52533 D. pteron
8	101	96.2	50	AAU96337	Aau96337 Der HMW-m
9	101	96.2	50	AAU96338	Aau96338 Der HMW-m
10	99	94.3	53	AAV52525	Aay52525 House dus
11	99	94.3	53	AAU96329	Aau96329 Der HMW-m
12	99	94.3	55	AAV52523	Aay52523 House dus
13	99	94.3	55	AAU96327	Aau96327 Der HMW-m
14	99	94.3	55	AAU96328	Aau96328 Der HMW-m
15	50	47.6	48	ABR68081	Abb68081 Drosophila
16	48	45.7	57	ABR58145	Abb58145 Lung can
17	48	45.7	117	ABG71824	Abg71824 Wild type
18	48	45.7	130	AAU98987	Aau98987 TNF relat
19	48	45.7	350	AAV00934	Aav00934 Human DR5
20	48	45.7	411	AAW76827	Aaw76827 Human TR6
21	48	45.7	411	AAW79083	Aaw79083 Human dea
22	48	45.7	411	AAW79261	Aaw79261 Tumour ne
23	48	45.7	411	AAW68410	Aaw68410 Human Apo
24	48	45.7	411	AAW63321	Aaw63321 Human Apo
25	48	45.7	411	AAW93576	Aaw93576 Human hAp

26	48	45.7	411	2	AAW93608	Aaw93608 Human x11
27	48	45.7	411	2	AAV00932	Aay00932 Human DR5
28	48	45.7	411	3	AAV55805	Aay55805 Human Apo
29	48	45.7	411	3	ABR29790	Abb29790 Human dea
30	48	45.7	411	3	ABR73443	Abb73443 Human Apo
31	48	45.7	411	4	ABR73442	Abb73442 Human Apo
32	48	45.7	411	4	AAU04038	Aau04038 Human apo
33	48	45.7	411	4	ABR48348	Abb48348 Human tum
34	48	45.7	411	5	ABR31490	Abb31490 Amino aci
35	48	45.7	411	5	AAU76947	Aau76947 Human Apo
36	48	45.7	411	5	ABR09602	Abb09602 Amino aci
37	48	45.7	411	5	AAV51080	Aay51080 Human DR5
38	48	45.7	411	6	ABU03488	Abu03488 Angiogene
39	48	45.7	411	6	ABR75645	Abb75645 Human Tan
40	48	45.7	411	6	ABR58626	Abb58626 Human can
41	48	45.7	411	6	ABP70904	Abp70904 Human FAM
42	48	45.7	411	6	AAO29541	Aao29541 Human TRA
43	48	45.7	411	6	ABG71903	Abg71903 Human TRA
44	48	45.7	411	6	ABG74383	Abg74383 Human TNF
45	48	45.7	411	6	ABR36243	Abb36243 Human TRA

ALIGNMENTS

RESULT 1  
AAV52513  
ID AAY52513 strand; peptide; 20 AA.  
XX  
AC AAY52513;  
XX  
DT 22-FEB-2000 (first entry)  
XX  
DE House dust mite allergen protein (map) A/B fragment map(3).  
XX  
KM Mite allergen protein; map; high molecular weight; HMW-map; allergy;  
KM house dust mite; IgE; immunoglobulin E; allergen; mapA; mapB;  
KM hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
KM canine; veterinary; antibody; vaccine; immunisation.  
XX  
OS Dermatophagoides farinae.  
XX  
PN WO954349-A2.  
XX  
PD 28-OCT-1999.  
XX  
PF 16-APR-1999; 99WO-US008524.  
XX  
PR 17-APR-1998; 98US-00062013.  
PR 13-MAY-1998; 98US-0085295P.  
PR 02-SEP-1998; 98US-0098905P.  
XX  
PA (HESK-) HESKA CORP.  
PI Mccall CA, Hunter SW, Weber ER;  
XX  
DR WPI; 2000-052700/04.  
XX  
PT Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
XX used to modify an animals' hypersensitivity to mite allergens.  
XX  
PS Claim 3; Page 69; 15pp; English.  
XX  
CC Sequences AAY52510-Y52522 represent proteolytic fragments of  
CC Dermatophagoides farinae high molecular weight mite allergen protein (HMW  
CC -map) composition. The HMW-map composition was isolated from a D. farinae  
CC homogenate by gel filtration, with each fraction being analysed for the  
CC presence of proteins that bound to IgE present in mite-allergic dog  
CC antisera. The HMW-map composition comprises mapA (a 109 kD protein) and  
CC mapB (98 kD). Mite allergenic proteins and peptides, and nucleic acids  
CC encoding them, may be used in therapeutic compositions to modify an  
CC animal's hypersensitivity reaction to mite allergens. Animals that may be  
CC treated include mammals and birds, especially felines, canines, equines,

CC humans, other pets, and work or domestic animals. The proteins or  
 CC fragments may also be used to diagnose allergies via a skin test. The  
 CC proteins and peptides can also be used to raise antibodies, which have a  
 CC variety of potential uses. For example, they can be used as vaccines to  
 CC passively immunise animals against dust mite hypersensitivity, as  
 CC positive controls in test kits and as tools to recover desired dust mite  
 CC allergens from a mixture of proteins

XX Sequence 20 AA;

Query Match 100.0%; Score 105; DB 3; Length 20;

Best Local Similarity 100.0%; Pred. No. 3.1e-09;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPAKMSPPGFIVGEGVLS 20  
 DB 1 DPAKMSPPGFIVGEGVLS 20

RESULT 2

AA52522

ID AA52522 standard; peptide; 20 AA.

XX AA52522;

DT 22-FEB-2000 (first entry)

DE House dust mite allergen protein (map) A/B fragment map(12).

XX Mite allergen protein; map; high molecular weight; HMW-map; allergy;

XX house dust mite; IGE; immunoglobulin E; allergen; mapA; mapB;

KM hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;

KM canine; veterinary; antibody; vaccine; immunisation.

XX Dermatophagoides farinae.

XX WO954349-A2.

PD 28-OCT-1999.

PF 16-APR-1999; 99WO-US008524.

XX 17-APR-1998; 98US-00062013.

PR 13-MAY-1998; 98US-0085295P.

PR 02-SEP-1998; 98US-0098909P.

XX (HESK-) HESKA CORP.

PI McCall CA, Hunter SW, Weber ER;

DR WPI; 2000-052700/04.

PT Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
 PT used to modify an animals' hypersensitivity to mite allergens.

XX Claim 3; Page 70; 154pp; English.

CC Sequences AA52520-Y52522 represent proteolytic fragments of

CC Dermatophagoides farinae high molecular weight mite allergen protein (HMW

CC -map) composition. The HMW-map composition was isolated from a D. farinae

CC homogenate by gel filtration, with each fraction being analysed for the

CC presence of proteins that bound to IGE present in mite-allergic dog

CC antisera. The HMW-map composition comprises mapA (a 109 kD protein) and

CC mapB (98 kD). Mite allergenic proteins and peptides, and nucleic acids

CC encoding them, may be used in therapeutic compositions to modify an

CC animal's hypersensitivity reaction to mite allergens. Animals that may be

CC treated include mammals and birds, especially felines, canines, equines,

CC humans, other pets, and work or domestic animals. The proteins or

CC fragments may also be used to diagnose allergies via a skin test. The

CC proteins and peptides can also be used to raise antibodies, which have a

CC variety of potential uses. For example, they can be used as vaccines to

CC passively immunise animals against dust mite hypersensitivity, as

CC positive controls in test kits and as tools to recover desired dust mite

CC allergens from a mixture of proteins

XX Sequence 20 AA;

Query Match 100.0%; Score 105; DB 3; Length 20;

Best Local Similarity 100.0%; Pred. No. 3.1e-09;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPAKMSPPGFIVGEGVLS 20  
 DB 1 DPAKMSPPGFIVGEGVLS 20

RESULT 3

AAU96317

ID AAU96317 standard; peptide; 20 AA.

XX AAU96317;

DT 15-JUL-2002 (first entry)

DE Der HMW-map polypeptide #4.

XX Der HMW-map; American house dust mite; anti-allergic; mite; IGE;

KM mite allergenic protein; immunoglobulin E; hypersensitivity;

XX immunocomplex formation.

XX Dermatophagoides farinae.

XX WO200222807-A2.

PD 21-MAR-2002.

PF 14-SEP-2001; 2001WO-US028730.

PR 14-SEP-2000; 2000US-00662293.

XX (HESK-) HESKA CORP.

PI McCall CA, Hunter SW, Weber ER;

DR WPI; 2002-351888/38.

PT New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HMW-map protein, useful as a vaccine for treating mite allergy.

XX Claim 12; Page 70; 161pp; English.

CC The invention relates to an isolated mite allergenic protein of

CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic

CC acid. The Der HMW-map protein is useful for eliciting an immune response

CC against Der HMW-map protein. The protein or a reagent comprising a non-

CC proteinaceous epitope is useful for identifying an animal (e.g., dog,

CC cat) susceptible to or having an allergic response to a mite. A

CC therapeutic composition is useful for desensitising a host animal to an

CC allergic response to a mite. The DNA and protein can be used in the

CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition

CC of immunoglobulin (Ig)E or Der HMW-map protein activity associated with a

CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting

CC binding of proteins to IGE, to prevent immunocomplex formation, thus

CC reducing hypersensitivity responses to mite allergens, and as vaccines

CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342

XX represent Der HMW-map polypeptides of the invention

XX Sequence 20 AA;

Query Match 100.0%; Score 105; DB 5; Length 20;

Best Local Similarity 100.0%; Pred. No. 3.1e-09;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPAKMSPPGFIVGEGVLS 20  
 DB 1 DPAKMSPPGFIVGEGVLS 20

```

RESULT 4
AAU96326 standard; peptide; 20 AA.
ID AAU96326 standard; peptide; 20 AA.
XX
AC AAU96326;
XX
DT 15-JUL-2002 (first entry)
XX
DE Der HMW-map polypeptide #13.
XX
XX Der HMW-map; American house dust mite; antiallergic; mite; IgE;
KM mite allergenic protein; immunoglobulin E; hypersensitivity;
XX immunocomplex formation.
XX
OS Dermatophagoides farinae.
XX
PN WO200222807-A2.
XX
PD 21-MAR-2002.
XX
PF 14-SEP-2001; 2001WO-US028730.
XX
PR 14-SEP-2000; 2000US-00662293.
XX
PA (HESK-) HESKA CORP.
XX
PI Mccall CA, Hunter SW, Weber ER;
XX
DR WPI; 2002-351888/38.
XX
PT New mite allergenic protein isolated from Dermatophagoides, designated
PT Der HMW-map protein, useful as a vaccine for treating mite allergy.
XX
PS Claim 12; Page 71; 161pp; English.
XX
XX The invention relates to an isolated mite allergenic protein of
CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic
CC acid. The Der HMW-map protein is useful for eliciting an immune response
CC against Der HMW-map protein. The protein or a reagent comprising a non-
CC proteinaceous epitope is useful for identifying an animal (e.g., dog,
CC cat) susceptible to or having an allergic response to a mite. A
CC therapeutic composition is useful for desensitizing a host animal to an
CC allergic response to a mite. The DNA and protein can be used in the
CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition
CC of immunoglobulin (Ig) E or Der HMW-map protein activity associated with a
CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting
CC binding of proteins to IgE, to prevent immunocomplex formation, thus
CC reducing hypersensitivity responses to mite allergens, and as vaccines
CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342
CC represent Der HMW-map polypeptides of the invention
XX
SQ Sequence 20 AA;
XX
Query Match 100.0%; Score 105; DB 5; Length 20;
Best Local Similarity 100.0%; Pred. No. 3, 1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DPAKMSPPGFIVGEGVLS 20
DB 1 DPAKMSPPGFIVGEGVLS 20
XX
RESULT 5
AAU96335 standard; protein; 490 AA.
ID AAU96335 standard; protein; 490 AA.
XX
AC AAU96335;
XX
DT 06-AUG-2003 (revised)
DT 22-FEB-2000 (first entry)
XX

```

```

DE D. pteronyssinus 98 kD mite allergen protein (map) Pderp98-490.
XX
XX Mite allergen protein; map; high molecular weight; HMW-map; allergy;
KM house dust mite; IgE; immunoglobulin E; allergen; mapB;
KM hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;
KM canine; veterinary; antibody; vaccine; immunisation.
XX
OS Dermatophagoides pteronyssinus.
XX
FH Key Location/Qualifiers
FT Modified-site 115..117
FT /note="Asn is N-glycosylated"
FT Modified-site 240..242
FT /note="Asn is N-glycosylated"
XX
PN WO9954349-A2.
XX
PD 28-OCT-1999.
XX
PF 16-APR-1999; 99WO-US008524.
XX
PR 17-APR-1998; 98US-00062013.
PR 13-MAY-1998; 98US-0085295P.
PR 02-SEP-1998; 98US-0098909P.
XX
PA (HESK-) HESKA CORP.
XX
PI Mccall CA, Hunter SW, Weber ER;
XX
DR WPI; 2000-052700/04.
DR N-PSDB; AA238589, AA238590.
XX
PT Novel high molecular weight Dermatophagoides nucleic acid polypeptides
PT used to modify an animals' hypersensitivity to mite allergens.
XX
XX Claim 3; Page 147-149; 154pp; English.
XX
XX This sequence represents Dermatophagoides pteronyssinus mite allergen
CC protein (map) Pderp98-490, the mature form of Pderp98-509. Pderp98-490
CC has a molecular weight of 98 kD, comprising 490 amino acids, and has a
CC high degree of homology with the D. farinae mature 98 kD allergen, mapB
CC (AA952525). Nucleic acid molecules encoding Pderp98-490 were isolated
CC from a D. pteronyssinus cDNA library by hybridisation with a probe
CC encoding the D. farinae high molecular weight map (HMW-map) composition.
CC Mite allergenic proteins and peptides, and nucleic acids encoding them,
CC may be used in therapeutic compositions to modify an animals'
CC hypersensitivity reaction to mite allergens. Animals that may be treated
CC include mammals and birds, especially felines, canines, equines, humans,
CC other pets, and work or domestic animals. The proteins or fragments may
CC also be used to diagnose allergies via a skin test. The proteins and
CC peptides can also be used to raise antibodies, which have a variety of
CC potential uses. For example, they can be used as vaccines to passively
CC immunise animals against dust mite hypersensitivity, as positive controls
CC in test kits and as tools to recover desired dust mite allergens from a
CC mixture of proteins. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ Sequence 490 AA;
XX
Query Match 96.2%; Score 101; DB 3; Length 490;
Best Local Similarity 95.0%; Pred. No. 4, 1e-07;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 DPAKMSPPGFIVGEGVLS 20
DB 279 DPAKMSPPGFIVGEGVLS 298
XX
RESULT 6
AAU96339 standard; protein; 490 AA.
ID AAU96339 standard; protein; 490 AA.
XX
AC AAU96339;
XX

```





```

XX OS Dermatophagoides farinae.
XX PN WO200222807-A2.
XX PD 21-MAR-2002.
XX PF 14-SEP-2001; 2001WO-US028730.
XX PR 14-SEP-2000; 2000US-00662293.
XX PA (HESK-) HESKA CORP.
XX PI McCall CA, Hunter SW, Weber ER;
XX DR WPI; 2002-351888/38.
XX DR N-PSDB; ABK69581.
XX PT New mite allergenic protein isolated from Dermatophagoides, designated
XX PT Der HMW-map protein, useful as a vaccine for treating mite allergy.
XX PS Claim 12; Page 134-136; 161pp; English.
XX CC The invention relates to an isolated mite allergenic protein of
XX CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic
XX CC acid. The Der HMW-map protein is useful for eliciting an immune response
XX CC against Der HMW-map protein. The protein or a reagent comprising a non-
XX CC proteinaceous epitope is useful for identifying an animal (e.g., dog,
XX CC cat) susceptible to or having an allergic response to a mite. A
XX CC therapeutic composition is useful for desensitizing a host animal to an
XX CC allergic response to a mite. The DNA and protein can be used in the
XX CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition
XX CC of immunoglobulin (Ig)E or Der HMW-map protein activity associated with a
XX CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting
XX CC binding of proteins to IgE, to prevent immunocomplex formation, thus
XX CC reducing hypersensitivity responses to mite allergens, and as vaccines
XX CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342
XX CC represent Der HMW-map polypeptides of the invention
SQ Sequence 509 AA;

Query Match          96.2%; Score 101, DB 5; Length 509;
Best Local Similarity 95.0%; Pred. No. 4.2e-07;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DPAKMSPPGFITGEGVLS 20
   |||||
DB 298 DPAKMSPPGFITGEGVLS 317

RESULT 9
AAU96338
ID AAU96338 standard; protein; 509 AA.
XX AC AAU96338;
XX DT 15-JUL-2002 (first entry)
XX DE Der HMW-map polypeptide #25.
XX KM Der HMW-map; American house dust mite; anti-allergic; mite; IgE;
XX KM mite allergenic protein; immunoglobulin E; hypersensitivity;
XX KM immunocomplex formation.
XX OS Dermatophagoides farinae.
XX PN WO200222807-A2.
XX PD 21-MAR-2002.
XX PF 14-SEP-2001; 2001WO-US028730.
XX PR 14-SEP-2000; 2000US-00662293.

```

```

XX PA (HESK-) HESKA CORP.
XX PI McCall CA, Hunter SW, Weber ER;
XX DR WPI; 2002-351888/38.
XX DR N-PSDB; ABK69583.
XX PT New mite allergenic protein isolated from Dermatophagoides, designated
XX PT Der HMW-map protein, useful as a vaccine for treating mite allergy.
XX PS Claim 12; Page 139-141; 161pp; English.
XX CC The invention relates to an isolated mite allergenic protein of
XX CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic
XX CC acid. The Der HMW-map protein is useful for eliciting an immune response
XX CC against Der HMW-map protein. The protein or a reagent comprising a non-
XX CC proteinaceous epitope is useful for identifying an animal (e.g., dog,
XX CC cat) susceptible to or having an allergic response to a mite. A
XX CC therapeutic composition is useful for desensitizing a host animal to an
XX CC allergic response to a mite. The DNA and protein can be used in the
XX CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition
XX CC of immunoglobulin (Ig)E or Der HMW-map protein activity associated with a
XX CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting
XX CC binding of proteins to IgE, to prevent immunocomplex formation, thus
XX CC reducing hypersensitivity responses to mite allergens, and as vaccines
XX CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342
XX CC represent Der HMW-map polypeptides of the invention
SQ Sequence 509 AA;

Query Match          96.2%; Score 101, DB 5; Length 509;
Best Local Similarity 95.0%; Pred. No. 4.2e-07;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DPAKMSPPGFITGEGVLS 20
   |||||
DB 298 DPAKMSPPGFITGEGVLS 317

RESULT 10
AAU52525
ID AAU52525 standard; protein; 536 AA.
XX AC AAU52525;
XX DT 22-FEB-2000 (first entry)
XX DE House dust mite (D. farinae) mite allergen protein (map) Pderf98-536.
XX KM Mite allergen protein; map; high molecular weight; HMW-map; allergy;
XX KM house dust mite; IgE; immunoglobulin E; allergen; map; map;
XX KM hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;
XX KM canine; veterinary; antibody; vaccine; immunisation.
XX OS Dermatophagoides farinae.
XX PN WO9954349-A2.
XX PD 28-OCT-1999.
XX PF 16-APR-1999; 99WO-US008524.
XX PR 17-APR-1998; 98US-00062013.
XX PR 13-MAY-1998; 98US-0085295P.
XX PR 02-SEP-1998; 98US-0098903P.
XX PA (HESK-) HESKA CORP.
XX PI McCall CA, Hunter SW, Weber ER;
XX DR WPI; 2000-052700/04.
XX DR N-PSDB; AAU38579, AAU38580.

```

XX Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
 PT used to modify an animals' hypersensitivity to mite allergens.  
 XX Claim 3; Page 125-127; 154pp; English.

XX This sequence represents Dermatophagoides farinae mite allergen protein  
 CC (map) Pderf98-555, the mature form of Pderf98-555 (AA52523). Pderf98-555  
 CC has a molecular weight of 98 kD, comprising 536 amino acids, and is a  
 CC component of the Dermatophagoides farinae high molecular weight mite  
 CC allergen protein (HMM-map) composition. The HMM-map composition was  
 CC isolated from a D. farinae homogenate by gel filtration, with each  
 CC fraction being analysed for the presence of proteins that bound to IgE  
 CC present in mite-allergic dog antisera. Mite allergenic proteins and  
 CC peptides, and nucleic acids encoding them, may be used in therapeutic  
 CC compositions to modify an animal's hypersensitivity reaction to mite  
 CC allergens. Animals that may be treated include mammals and birds,  
 CC especially felines, canines, equines, humans, other pets, and work or  
 CC domestic animals. The proteins or fragments may also be used to diagnose  
 CC allergies via a skin test. The proteins and peptides can also be used to  
 CC raise antibodies, which have a variety of potential uses. For example,  
 CC they can be used as vaccines to passively immunise animals against dust  
 CC mite hypersensitivity, as positive controls in test kits and as tools to  
 CC recover desired dust mite allergens from a mixture of proteins

XX Sequence 536 AA;

Query Match 94.3%; Score 99; DB 3; Length 536;  
 Best Local Similarity 95.0%; Pred. No. 9.1e-07;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DPAKMSPPGFVSGSEGYLS 20  
 |||||  
 Db 279 DPAKMSPPGFVSGSEGYLS 298

RESULT 11  
 AAU96329  
 ID AAU96329 standard; protein; 536 AA.

XX AAU96329;  
 AC  
 XX 15-JUL-2002 (first entry)  
 DT  
 XX  
 DE Der HMM-map polypeptide #16.

XX Der HMM-map; American house dust mite; antiallergic; mite; IgE;  
 KM mite allergenic protein; immunoglobulin E; hypersensitivity;  
 KM immunocomplex formation.  
 XX  
 OS Dermatophagoides farinae.  
 XX  
 PN WO200222807-A2.  
 PD 21-MAR-2002.  
 XX  
 PF 14-SEP-2001; 2001WO-US028730.  
 XX  
 PR 14-SEP-2000; 2000US-00662293.  
 XX  
 PA (HESK-) HESKA CORP.  
 XX  
 PI McCall CA, Hunter SW, Weber ER;  
 XX  
 DR WPI; 2002-351888/38.  
 DR N-PsDB; ABK69575.  
 XX  
 PT New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HMM-map protein, useful as a vaccine for treating mite allergy.  
 PS Claim 12; Page 125-127; 161pp; English.  
 XX The invention relates to an isolated mite allergenic protein of

CC Dermatophagoides, designated Der HMM-map protein, and its related nucleic  
 CC acid. The Der HMM-map protein is useful for eliciting an immune response  
 CC against Der HMM-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitising a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMM-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig)E or Der HMM-map protein actively associated with a  
 CC disease. Antibodies that bind to Der HMM-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HMM-map polypeptides of the invention

XX Sequence 536 AA;

Query Match 94.3%; Score 99; DB 5; Length 536;  
 Best Local Similarity 95.0%; Pred. No. 9.1e-07;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DPAKMSPPGFVSGSEGYLS 20  
 |||||  
 Db 279 DPAKMSPPGFVSGSEGYLS 298

RESULT 12  
 AA52523  
 ID AA52523 standard; protein; 555 AA.

XX AA52523;  
 AC  
 XX 22-FEB-2000 (first entry)  
 DT  
 XX  
 DE House dust mite (D. farinae) mite allergen protein (map) Pderf98-555.

XX Mite allergen protein; map; high molecular weight; HMM-map; allergy;  
 KM house dust mite; IgE; immunoglobulin E; allergen; mapA; mapB;  
 KM hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
 KM canine; veterinary; antibody; vaccine; immunisation.  
 XX  
 OS Dermatophagoides farinae.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..19 "Signal peptide"  
 FT Protein 20..555  
 FT "note= \"Mature Pderf98-555\"  
 XX  
 PN WO954349-A2.  
 PD 28-OCT-1999.  
 XX  
 PF 16-APR-1999; 99WO-US008524.  
 XX  
 PR 17-APR-1998; 98US-00062013.  
 PR 13-MAY-1998; 98US-0085295P.  
 PR 02-SEP-1998; 98US-0098909P.  
 XX  
 PA (HESK-) HESKA CORP.  
 XX  
 PI McCall CA, Hunter SW, Weber ER;  
 XX  
 DR WPI; 2000-052700/04.  
 DR N-PsDB; AA238575, AA238576, AA238577, AA238578.  
 XX  
 PT Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
 PT used to modify an animals' hypersensitivity to mite allergens.  
 PS Claim 3; Page 111-113; 154pp; English.  
 XX This sequence represents Dermatophagoides farinae mite allergen protein  
 CC (map) Pderf98-555. Pderf98-555 has a molecular weight of 98 kD,

comprising 555 amino acids, and is a component of the Dermatophagoides farinae high molecular weight mite allergen protein (HWM-map) composition. The HWM-map composition was isolated from a D. farinae homogenate by gel filtration, with each fraction being analysed for the presence of proteins that bound to IgE present in mite-allergic dog antiserum. Mite allergenic proteins and peptides, and nucleic acids encoding them, may be used in therapeutic compositions to modify an animal's hypersensitivity reaction to mite allergens. Animals that may be treated include mammals and birds, especially felines, canines, equines, humans, other pets, and work or domestic animals. The proteins or fragments may also be used to diagnose allergies via a skin test. The proteins and peptides can also be used to raise antibodies, which have a variety of potential uses. For example, they can be used as vaccines to passively immunise animals against dust mite hypersensitivity, as positive controls in test kits and as tools to recover desired dust mite allergens from a mixture of proteins

Query Match 94.3%; Score 99; DB 3; Length 555;  
Best Local Similarity 95.0%; Pred. No. 9.5e-07;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 DPAKMSPPGFIVGEGVLS 20  
298 DPAKMSPPGFIVGEGVLS 317

RESULT 13  
AAU96327.  
ID AAU96327 standard; protein; 555 AA.  
AC AAU96327;  
XX  
DT 15-JUL-2002 (first entry)  
XX  
DE Der HWM-map polypeptide #14.  
XX  
XX Der HWM-map; American house dust mite; antiallergic; mite; IgE;  
KM mite allergenic protein; immunoglobulin E; hypersensitivity;  
XX immunocomplex formation.  
XX  
OS Dermatophagoides farinae.  
XX  
PN WO200222807-A2.  
XX  
PD 21-MAR-2002.  
XX  
PF 14-SEP-2001; 2001WO-US028730.  
XX  
PR 14-SEP-2000; 2000US-00662293.  
XX  
PA (HESK-) HESKA CORP.  
XX  
PI Mccall CA, Hunter SM, Weber ER;  
XX  
XX WPI; 2002-351888/38.  
DR N-PSDB; ABR69571.  
XX  
XX  
PT New mite allergenic protein isolated from Dermatophagoides, designated  
XX Der HWM-map protein, useful as a vaccine for treating mite allergy.  
XX  
PS Claim 12; Page 114-116; 161pp; English.  
XX  
XX The invention relates to an isolated mite allergenic protein of  
CC Dermatophagoides, designated Der HWM-map protein, and its related nucleic  
CC acid. The Der HWM-map protein is useful for eliciting an immune response  
CC against Der HWM-map protein. The protein or a reagent comprising a non-  
CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
CC cat) susceptible to or having an allergic response to a mite. A  
CC therapeutic composition is useful for desensitising a host animal to an  
CC allergic response to a mite. The DNA and protein can be used in the  
CC detection of anti-Der HWM-map antibodies in animal fluids, and inhibition

of immunoglobulin (Ig)E or Der HWM-map protein activity associated with a disease. Antibodies that bind to Der HWM-map are useful for inhibiting CC binding of proteins to IgE, to prevent immunocomplex formation, thus CC reducing hypersensitivity responses to mite allergens, and as vaccines CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342 CC represent Der HWM-map polypeptides of the invention

Query Match 94.3%; Score 99; DB 5; Length 555;  
Best Local Similarity 95.0%; Pred. No. 9.5e-07;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 DPAKMSPPGFIVGEGVLS 20  
298 DPAKMSPPGFIVGEGVLS 317

RESULT 14  
AAU96328  
ID AAU96328 standard; protein; 555 AA.  
AC AAU96328;  
XX  
DT 15-JUL-2002 (first entry)  
XX  
DE Der HWM-map polypeptide #15.  
XX  
XX Der HWM-map; American house dust mite; antiallergic; mite; IgE;  
KM mite allergenic protein; immunoglobulin E; hypersensitivity;  
XX immunocomplex formation.  
XX  
OS Dermatophagoides farinae.  
XX  
PN WO200222807-A2.  
XX  
PD 21-MAR-2002.  
XX  
PF 14-SEP-2001; 2001WO-US028730.  
XX  
PR 14-SEP-2000; 2000US-00662293.  
XX  
PA (HESK-) HESKA CORP.  
XX  
PI Mccall CA, Hunter SM, Weber ER;  
XX  
XX WPI; 2002-351888/38.  
DR N-PSDB; ABR69573.  
XX  
XX  
PT New mite allergenic protein isolated from Dermatophagoides, designated  
XX Der HWM-map protein, useful as a vaccine for treating mite allergy.  
XX  
PS Claim 12; Page 120-122; 161pp; English.  
XX  
XX The invention relates to an isolated mite allergenic protein of  
CC Dermatophagoides, designated Der HWM-map protein, and its related nucleic  
CC acid. The Der HWM-map protein is useful for eliciting an immune response  
CC against Der HWM-map protein. The protein or a reagent comprising a non-  
CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
CC cat) susceptible to or having an allergic response to a mite. A  
CC therapeutic composition is useful for desensitising a host animal to an  
CC allergic response to a mite. The DNA and protein can be used in the  
CC detection of anti-Der HWM-map antibodies in animal fluids, and inhibition  
CC of immunoglobulin (Ig)E or Der HWM-map protein activity associated with a  
CC disease. Antibodies that bind to Der HWM-map are useful for inhibiting  
CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
CC reducing hypersensitivity responses to mite allergens, and as vaccines  
CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
CC represent Der HWM-map polypeptides of the invention

Query Match 94.3%; Score 99; DB 5; Length 555;  
Sequence 555 AA;

Best Local Similarity 95.0%; Pred. No. 9.5e-07;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 DPAKMSPPGFIIVGEGVLS 20  
Db 298 DPAKMSPPGFIIVGEGVLS 317

RESULT 15

ABB68081  
ID ABB68081 standard; protein; 488 AA.

AC ABB68081;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 31035.

KM Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US009231.

PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656660/75.

DR N-PSDB; ABL12184.

PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signaling and cell-cell interactions.

PS Disclosure; SEQ ID NO 31035; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signaling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL10511); expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (AB57737-AB572072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

CC Sequence 488 AA;

Query Match 47.6%; Score 50; DB 4; Length 488;  
Best Local Similarity 40.0%; Pred. No. 33;  
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 1 DPAKMSPPGFIIVGEGVLS 20  
Db 407 DPAKMSPPGFIIVGEGVLS 426

Search completed: March 22, 2004, 06:51:36  
Job time : 8.0816 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: March 22, 2004, 06:59:34 ; Search time 4.0916 Seconds  
(without alignments)  
1265.926 Million cell updates/sec

Title: US-09-662-293-4

Sequence: 1 DPAKMSPPGFIVGEEGVLS 20

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1049977 seqs, 258955339 residues

Total number of hits satisfying chosen parameters: 1049977

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: Published Applications AA:\*  
2: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB.pep:\*  
3: /cgn2\_6/ptodata/2/pubppaa/PCF\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep:\*  
5: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/2/pubppaa/PCFUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubppaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubppaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubppaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubppaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubppaa/US10C\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	100.0	20	14	US-10-218-743-4
2	105	100.0	20	14	US-10-218-743-13
3	101	96.2	490	14	US-10-218-743-41
4	101	96.2	509	14	US-10-218-743-35
5	101	96.2	509	14	US-10-218-743-38
6	99	94.3	536	14	US-10-218-743-21
7	99	94.3	555	14	US-10-218-743-15
8	99	94.3	555	14	US-10-218-743-18
9	50	47.6	143	12	US-10-424-599-238601
10	49	46.7	69	12	US-10-424-599-181344
11	48	45.7	57	9	US-09-925-302-483
12	48	45.7	350	13	US-10-067-615-6
13	48	45.7	350	13	US-10-076-754-6
14	48	45.7	350	13	US-10-076-773-6
15	48	45.7	411	8	US-08-916-625B-2

16	48	45.7	411	9	US-09-828-739-2	Sequence 2, Appl1
17 <td>48<td>45.7<th>411</th><th>9</th><th>US-09-757-421-4</th><th>Sequence 4, Appl1</th></td></td>	48 <td>45.7<th>411</th><th>9</th><th>US-09-757-421-4</th><th>Sequence 4, Appl1</th></td>	45.7 <th>411</th> <th>9</th> <th>US-09-757-421-4</th> <th>Sequence 4, Appl1</th>	411	9	US-09-757-421-4	Sequence 4, Appl1
18 <td>48<td>45.7<th>411</th><th>9</th><th>US-09-874-138-2</th><th>Sequence 2, Appl1</th></td></td>	48 <td>45.7<th>411</th><th>9</th><th>US-09-874-138-2</th><th>Sequence 2, Appl1</th></td>	45.7 <th>411</th> <th>9</th> <th>US-09-874-138-2</th> <th>Sequence 2, Appl1</th>	411	9	US-09-874-138-2	Sequence 2, Appl1
19 <td>48<td>45.7<th>411</th><th>9</th><th>US-09-867-879-11</th><th>Sequence 11, Appl1</th></td></td>	48 <td>45.7<th>411</th><th>9</th><th>US-09-867-879-11</th><th>Sequence 11, Appl1</th></td>	45.7 <th>411</th> <th>9</th> <th>US-09-867-879-11</th> <th>Sequence 11, Appl1</th>	411	9	US-09-867-879-11	Sequence 11, Appl1
20 <td>48<td>45.7<th>411</th><th>9</th><th>US-09-811-088-6</th><th>Sequence 6, Appl1</th></td></td>	48 <td>45.7<th>411</th><th>9</th><th>US-09-811-088-6</th><th>Sequence 6, Appl1</th></td>	45.7 <th>411</th> <th>9</th> <th>US-09-811-088-6</th> <th>Sequence 6, Appl1</th>	411	9	US-09-811-088-6	Sequence 6, Appl1
21 <td>48<td>45.7<th>411</th><th>9</th><th>US-09-992-964-11</th><th>Sequence 11, Appl1</th></td></td>	48 <td>45.7<th>411</th><th>9</th><th>US-09-992-964-11</th><th>Sequence 11, Appl1</th></td>	45.7 <th>411</th> <th>9</th> <th>US-09-992-964-11</th> <th>Sequence 11, Appl1</th>	411	9	US-09-992-964-11	Sequence 11, Appl1
22 <td>48<td>45.7<th>411</th><th>12</th><th>US-10-211-462-67</th><th>Sequence 67, Appl1</th></td></td>	48 <td>45.7<th>411</th><th>12</th><th>US-10-211-462-67</th><th>Sequence 67, Appl1</th></td>	45.7 <th>411</th> <th>12</th> <th>US-10-211-462-67</th> <th>Sequence 67, Appl1</th>	411	12	US-10-211-462-67	Sequence 67, Appl1
23 <td>48<td>45.7<th>411</th><th>13</th><th>US-10-039-785-3</th><th>Sequence 3, Appl1</th></td></td>	48 <td>45.7<th>411</th><th>13</th><th>US-10-039-785-3</th><th>Sequence 3, Appl1</th></td>	45.7 <th>411</th> <th>13</th> <th>US-10-039-785-3</th> <th>Sequence 3, Appl1</th>	411	13	US-10-039-785-3	Sequence 3, Appl1
24 <td>48<td>45.7<th>411</th><th>13</th><th>US-10-005-842-2</th><th>Sequence 2, Appl1</th></td></td>	48 <td>45.7<th>411</th><th>13</th><th>US-10-005-842-2</th><th>Sequence 2, Appl1</th></td>	45.7 <th>411</th> <th>13</th> <th>US-10-005-842-2</th> <th>Sequence 2, Appl1</th>	411	13	US-10-005-842-2	Sequence 2, Appl1
25 <td>48<td>45.7<th>411</th><th>13</th><th>US-10-067-615-2</th><th>Sequence 2, Appl1</th></td></td>	48 <td>45.7<th>411</th><th>13</th><th>US-10-067-615-2</th><th>Sequence 2, Appl1</th></td>	45.7 <th>411</th> <th>13</th> <th>US-10-067-615-2</th> <th>Sequence 2, Appl1</th>	411	13	US-10-067-615-2	Sequence 2, Appl1
26 <td>48<td>45.7<th>411</th><th>13</th><th>US-10-052-758-1</th><th>Sequence 2, Appl1</th></td></td>	48 <td>45.7<th>411</th><th>13</th><th>US-10-052-758-1</th><th>Sequence 2, Appl1</th></td>	45.7 <th>411</th> <th>13</th> <th>US-10-052-758-1</th> <th>Sequence 2, Appl1</th>	411	13	US-10-052-758-1	Sequence 2, Appl1
27 <td>48<td>45.7<th>411</th><th>13</th><th>US-10-076-754-2</th><th>Sequence 1, Appl1</th></td></td>	48 <td>45.7<th>411</th><th>13</th><th>US-10-076-754-2</th><th>Sequence 1, Appl1</th></td>	45.7 <th>411</th> <th>13</th> <th>US-10-076-754-2</th> <th>Sequence 1, Appl1</th>	411	13	US-10-076-754-2	Sequence 1, Appl1
28 <td>48<td>45.7<th>411</th><th>13</th><th>US-10-076-773-2</th><th>Sequence 2, Appl1</th></td></td>	48 <td>45.7<th>411</th><th>13</th><th>US-10-076-773-2</th><th>Sequence 2, Appl1</th></td>	45.7 <th>411</th> <th>13</th> <th>US-10-076-773-2</th> <th>Sequence 2, Appl1</th>	411	13	US-10-076-773-2	Sequence 2, Appl1
29 <td>48<td>45.7<th>411</th><th>14</th><th>US-10-207-655-194</th><th>Sequence 194, App</th></td></td>	48 <td>45.7<th>411</th><th>14</th><th>US-10-207-655-194</th><th>Sequence 194, App</th></td>	45.7 <th>411</th> <th>14</th> <th>US-10-207-655-194</th> <th>Sequence 194, App</th>	411	14	US-10-207-655-194	Sequence 194, App
30 <td>48<td>45.7<th>411</th><th>14</th><th>US-10-314-410-6</th><th>Sequence 6, Appl1</th></td></td>	48 <td>45.7<th>411</th><th>14</th><th>US-10-314-410-6</th><th>Sequence 6, Appl1</th></td>	45.7 <th>411</th> <th>14</th> <th>US-10-314-410-6</th> <th>Sequence 6, Appl1</th>	411	14	US-10-314-410-6	Sequence 6, Appl1
31 <td>48<td>45.7<th>411</th><th>14</th><th>US-10-242-383-11</th><th>Sequence 11, Appl1</th></td></td>	48 <td>45.7<th>411</th><th>14</th><th>US-10-242-383-11</th><th>Sequence 11, Appl1</th></td>	45.7 <th>411</th> <th>14</th> <th>US-10-242-383-11</th> <th>Sequence 11, Appl1</th>	411	14	US-10-242-383-11	Sequence 11, Appl1
32 <td>48<td>45.7<th>411</th><th>14</th><th>US-10-288-917-1</th><th>Sequence 1, Appl1</th></td></td>	48 <td>45.7<th>411</th><th>14</th><th>US-10-288-917-1</th><th>Sequence 1, Appl1</th></td>	45.7 <th>411</th> <th>14</th> <th>US-10-288-917-1</th> <th>Sequence 1, Appl1</th>	411	14	US-10-288-917-1	Sequence 1, Appl1
33 <td>48<td>45.7<th>411</th><th>14</th><th>US-10-322-673-3</th><th>Sequence 3, Appl1</th></td></td>	48 <td>45.7<th>411</th><th>14</th><th>US-10-322-673-3</th><th>Sequence 3, Appl1</th></td>	45.7 <th>411</th> <th>14</th> <th>US-10-322-673-3</th> <th>Sequence 3, Appl1</th>	411	14	US-10-322-673-3	Sequence 3, Appl1
34 <td>48<td>45.7<th>411</th><th>14</th><th>US-10-139-785-3</th><th>Sequence 3, Appl1</th></td></td>	48 <td>45.7<th>411</th><th>14</th><th>US-10-139-785-3</th><th>Sequence 3, Appl1</th></td>	45.7 <th>411</th> <th>14</th> <th>US-10-139-785-3</th> <th>Sequence 3, Appl1</th>	411	14	US-10-139-785-3	Sequence 3, Appl1
35 <td>48<td>45.7<th>411</th><th>15</th><th>US-10-297-486-3</th><th>Sequence 3, Appl1</th></td></td>	48 <td>45.7<th>411</th><th>15</th><th>US-10-297-486-3</th><th>Sequence 3, Appl1</th></td>	45.7 <th>411</th> <th>15</th> <th>US-10-297-486-3</th> <th>Sequence 3, Appl1</th>	411	15	US-10-297-486-3	Sequence 3, Appl1
36 <td>48<td>45.7<th>411</th><th>15</th><th>US-10-335-712-3</th><th>Sequence 3, Appl1</th></td></td>	48 <td>45.7<th>411</th><th>15</th><th>US-10-335-712-3</th><th>Sequence 3, Appl1</th></td>	45.7 <th>411</th> <th>15</th> <th>US-10-335-712-3</th> <th>Sequence 3, Appl1</th>	411	15	US-10-335-712-3	Sequence 3, Appl1
37 <td>48<td>45.7<th>411</th><th>15</th><th>US-10-423-448-1</th><th>Sequence 1, Appl1</th></td></td>	48 <td>45.7<th>411</th><th>15</th><th>US-10-423-448-1</th><th>Sequence 1, Appl1</th></td>	45.7 <th>411</th> <th>15</th> <th>US-10-423-448-1</th> <th>Sequence 1, Appl1</th>	411	15	US-10-423-448-1	Sequence 1, Appl1
38 <td>48<td>45.7<th>440</th><th>9</th><th>US-09-757-421-2</th><th>Sequence 2, Appl1</th></td></td>	48 <td>45.7<th>440</th><th>9</th><th>US-09-757-421-2</th><th>Sequence 2, Appl1</th></td>	45.7 <th>440</th> <th>9</th> <th>US-09-757-421-2</th> <th>Sequence 2, Appl1</th>	440	9	US-09-757-421-2	Sequence 2, Appl1
39 <td>48<td>45.7<th>440</th><th>9</th><th>US-09-811-088-4</th><th>Sequence 4, Appl1</th></td></td>	48 <td>45.7<th>440</th><th>9</th><th>US-09-811-088-4</th><th>Sequence 4, Appl1</th></td>	45.7 <th>440</th> <th>9</th> <th>US-09-811-088-4</th> <th>Sequence 4, Appl1</th>	440	9	US-09-811-088-4	Sequence 4, Appl1
40 <td>48<td>45.7<th>440</th><th>14</th><th>US-10-314-410-4</th><th>Sequence 4, Appl1</th></td></td>	48 <td>45.7<th>440</th><th>14</th><th>US-10-314-410-4</th><th>Sequence 4, Appl1</th></td>	45.7 <th>440</th> <th>14</th> <th>US-10-314-410-4</th> <th>Sequence 4, Appl1</th>	440	14	US-10-314-410-4	Sequence 4, Appl1
41 <td>48<td>45.7<th>440</th><th>15</th><th>US-10-335-712-3</th><th>Sequence 4, Appl1</th></td></td>	48 <td>45.7<th>440</th><th>15</th><th>US-10-335-712-3</th><th>Sequence 4, Appl1</th></td>	45.7 <th>440</th> <th>15</th> <th>US-10-335-712-3</th> <th>Sequence 4, Appl1</th>	440	15	US-10-335-712-3	Sequence 4, Appl1
42 <td>46<td>43.8<th>34</th><th>12</th><th>US-10-424-599-169740</th><th>Sequence 3, Appl1</th></td></td>	46 <td>43.8<th>34</th><th>12</th><th>US-10-424-599-169740</th><th>Sequence 3, Appl1</th></td>	43.8 <th>34</th> <th>12</th> <th>US-10-424-599-169740</th> <th>Sequence 3, Appl1</th>	34	12	US-10-424-599-169740	Sequence 3, Appl1
43 <td>46<td>43.8<th>135</th><th>12</th><th>US-10-424-599-267787</th><th>Sequence 169740,</th></td></td>	46 <td>43.8<th>135</th><th>12</th><th>US-10-424-599-267787</th><th>Sequence 169740,</th></td>	43.8 <th>135</th> <th>12</th> <th>US-10-424-599-267787</th> <th>Sequence 169740,</th>	135	12	US-10-424-599-267787	Sequence 169740,
44 <td>45<td>42.9<th>123</th><th>12</th><th>US-10-425-114-57295</th><th>Sequence 267787,</th></td></td>	45 <td>42.9<th>123</th><th>12</th><th>US-10-425-114-57295</th><th>Sequence 267787,</th></td>	42.9 <th>123</th> <th>12</th> <th>US-10-425-114-57295</th> <th>Sequence 267787,</th>	123	12	US-10-425-114-57295	Sequence 267787,
45 <td>45<td>42.9<th>303</th><th>9</th><th>US-09-738-626-6409</th><th>Sequence 57295, A</th></td></td>	45 <td>42.9<th>303</th><th>9</th><th>US-09-738-626-6409</th><th>Sequence 57295, A</th></td>	42.9 <th>303</th> <th>9</th> <th>US-09-738-626-6409</th> <th>Sequence 57295, A</th>	303	9	US-09-738-626-6409	Sequence 57295, A
						Sequence 6409, Ap

#### ALIGNMENTS

RESULT 1  
US-10-218-743-4  
Sequence 4, Application US/10218743  
Publication No. US20030096779A1  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley M.  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
TITLE OF INVENTION: AND USES THEREOF  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/10/218,743  
CURRENT FILING DATE: 2002-08-13  
PRIOR APPLICATION NUMBER: US/09/292,225  
PRIOR FILING DATE: 1999-04-15  
PRIOR APPLICATION NUMBER: 60/098,909  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/085,295  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/098,565  
PRIOR FILING DATE: 1998-04-17  
PRIOR APPLICATION NUMBER: 09/062,013  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4  
LENGTH: 20  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-10-218-743-4  
Query Match 100.0%; Score 105; DB 14; Length 20;  
Best Local Similarity 100.0%; Pred. No. 7.6e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 1 DPAKMSPPGFIVGEEGVLS 20  
DB 1 DPAKMSPPGFIVGEEGVLS 20

```
RESULT 2
US-10-218-743-13
; Sequence 13, Application US/10218743
; Publication No. US20030096779A1
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/10/218,743
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US/09/292,225
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: 60/098,909
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/085,295
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/098,565
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: 09/062,013
; PRIOR FILING DATE: 1998-04-17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-10-218-743-13

Query Match      100.0%; Score 105; DB 14; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY      1 DPAKMSPPGFTVGEVGLS 20
        |||||
Db      1 DPAKMSPPGFTVGEVGLS 20

RESULT 3
US-10-218-743-41
; Sequence 41, Application US/10218743
; Publication No. US20030096779A1
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/10/218,743
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US/09/292,225
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: 60/098,909
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/085,295
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/098,565
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: 09/062,013
; PRIOR FILING DATE: 1998-04-17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41
; LENGTH: 490
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-10-218-743-41
```

```
Query Match      96.2%; Score 101; DB 14; Length 490;
Best Local Similarity 95.0%; Pred. No. 9.4e-07;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CY      1 DPAKMSPPGFTVGEVGLS 20
        |||||
Db      279 DPAKMSPPGFTVGEVGLS 298

RESULT 4
US-10-218-743-35
; Sequence 35, Application US/10218743
; Publication No. US20030096779A1
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/10/218,743
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US/09/292,225
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: 60/098,909
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/085,295
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/098,565
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: 09/062,013
; PRIOR FILING DATE: 1998-04-17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-10-218-743-35

Query Match      96.2%; Score 101; DB 14; Length 509;
Best Local Similarity 95.0%; Pred. No. 9.8e-07;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CY      1 DPAKMSPPGFTVGEVGLS 20
        |||||
Db      298 DPAKMSPPGFTVGEVGLS 317

RESULT 5
US-10-218-743-38
; Sequence 38, Application US/10218743
; Publication No. US20030096779A1
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/10/218,743
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US/09/292,225
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: 60/098,909
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/085,295
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/098,565
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: 09/062,013
; PRIOR FILING DATE: 1998-04-17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 490
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-10-218-743-38
```

SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 38  
LENGTH: 509  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-10-218-743-38

Query Match 96.2%; Score 101; DB 14; Length 509;  
Best Local Similarity 95.0%; Pred. No. 9.8e-07;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DPAKMSPPGFIVGEGVLS 20  
|||||  
Db 298 DPAKMSPPGFIVGEGVLS 317

RESULT 6  
US-10-218-743-21  
Sequence 21, Application US/10218743  
Publication No. US20030096779A1  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley Wu  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/10/218,743  
CURRENT FILING DATE: 2002-08-13  
PRIOR APPLICATION NUMBER: US/09/292,225  
PRIOR FILING DATE: 1999-04-15  
PRIOR APPLICATION NUMBER: 60/098,909  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/085,295  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/098,565  
PRIOR FILING DATE: 1998-04-17  
PRIOR APPLICATION NUMBER: 09/062,013  
PRIOR FILING DATE: 1998-04-17  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 21  
LENGTH: 536  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-10-218-743-21

Query Match 94.3%; Score 99; DB 14; Length 536;  
Best Local Similarity 95.0%; Pred. No. 2.1e-06;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DPAKMSPPGFIVGEGVLS 20  
|||||  
Db 279 DPAKMSPPGFIVGEGVLS 298

RESULT 7  
US-10-218-743-15  
Sequence 15, Application US/10218743  
Publication No. US20030096779A1  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley Wu  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/10/218,743  
CURRENT FILING DATE: 2002-08-13  
PRIOR APPLICATION NUMBER: US/09/292,225  
PRIOR FILING DATE: 1999-04-15  
PRIOR APPLICATION NUMBER: 60/098,909  
PRIOR FILING DATE: 1998-09-02

PRIOR APPLICATION NUMBER: 60/085,295  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/098,565  
PRIOR FILING DATE: 1998-04-17  
PRIOR APPLICATION NUMBER: 09/062,013  
PRIOR FILING DATE: 1998-04-17  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 15  
LENGTH: 555  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-10-218-743-15

Query Match 94.3%; Score 99; DB 14; Length 555;  
Best Local Similarity 95.0%; Pred. No. 2.1e-06;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DPAKMSPPGFIVGEGVLS 20  
|||||  
Db 298 DPAKMSPPGFIVGEGVLS 317

RESULT 8  
US-10-218-743-18  
Sequence 18, Application US/10218743  
Publication No. US20030096779A1  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley Wu  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/10/218,743  
CURRENT FILING DATE: 2002-08-13  
PRIOR APPLICATION NUMBER: US/09/292,225  
PRIOR FILING DATE: 1999-04-15  
PRIOR APPLICATION NUMBER: 60/098,909  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/085,295  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/098,565  
PRIOR FILING DATE: 1998-04-17  
PRIOR APPLICATION NUMBER: 09/062,013  
PRIOR FILING DATE: 1998-04-17  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 18  
LENGTH: 555  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-10-218-743-18

Query Match 94.3%; Score 99; DB 14; Length 555;  
Best Local Similarity 95.0%; Pred. No. 2.1e-06;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DPAKMSPPGFIVGEGVLS 20  
|||||  
Db 298 DPAKMSPPGFIVGEGVLS 317

RESULT 9  
US-10-424-599-238601  
Sequence 238601, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J  
APPLICANT: Kovalic, David K  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 238601  
LENGTH: 143  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_57482C.1.pep  
US-10-424-599-238601

Query Match 47.6%; Score 50; DB 12; Length 143;  
Best Local Similarity 61.1%; Pred. No. 11;  
Matches 11; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

Qy 2 PAKGSPPGF--IVGREG 17  
Db 121 PAKGSDPPKFPNVGREG 138

RESULT 10  
US-10-424-599-181344  
Sequence 181344, Application US/10424559  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 181344  
LENGTH: 69  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_134769C.1.pep  
US-10-424-599-181344

Query Match 46.7%; Score 49; DB 12; Length 69;  
Best Local Similarity 72.7%; Pred. No. 6.8;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 PAKGSPPGF 12  
Db 31 PAKGSPPGF 41

RESULT 11  
US-09-925-302-483  
Sequence 483, Application US/09925302  
Patent No. US2002004941A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA104  
CURRENT APPLICATION NUMBER: US/09/925,302  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05918  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 896  
SOFTWARE: Patent Ver. 2.0  
SEQ ID NO 483  
LENGTH: 57  
TYPE: PRT

ORGANISM: Homo sapiens  
US-09-925-302-483

Query Match 45.7%; Score 48; DB 9; Length 57;  
Best Local Similarity 57.1%; Pred. No. 7.8;  
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 6 MSPPGFIVGEGVL 19  
Db 9 LSPGPMGVGEGVL 22

RESULT 12  
US-10-067-615-6  
Sequence 6, Application US/10067615  
Publication No. US20020115154A1  
GENERAL INFORMATION:  
APPLICANT: Alnemri, Emad S.  
TITLE OF INVENTION: NOVEL TRAIL RECEPTORS, NUCLEIC ACIDS ENCODING SAME, AND  
FILE REFERENCE: 480140.432D1  
CURRENT APPLICATION NUMBER: US/10/067,615  
CURRENT FILING DATE: 2002-02-04  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: Patent Ver. 2.0  
SEQ ID NO 6  
LENGTH: 350  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-067-615-6

Query Match 45.7%; Score 48; DB 13; Length 350;  
Best Local Similarity 43.8%; Pred. No. 55;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 2 PAKGSPPGFIVGEG 17  
Db 76 PSEGLCPGHHISEDG 91

RESULT 13  
US-10-076-754-6  
Sequence 6, Application US/10076754  
Publication No. US20020161195A1  
GENERAL INFORMATION:  
APPLICANT: Alnemri, Emad S.  
TITLE OF INVENTION: NOVEL TRAIL RECEPTORS, NUCLEIC ACIDS ENCODING SAME, AND  
FILE REFERENCE: 480140.432D3  
CURRENT APPLICATION NUMBER: US/10/076,754  
CURRENT FILING DATE: 2002-02-12  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: Patent Ver. 2.0  
SEQ ID NO 6  
LENGTH: 350  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-076-754-6

Query Match 45.7%; Score 48; DB 13; Length 350;  
Best Local Similarity 43.8%; Pred. No. 55;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 2 PAKGSPPGFIVGEG 17  
Db 76 PSEGLCPGHHISEDG 91

RESULT 14  
US-10-076-773-6  
Sequence 6, Application US/10076773  
Publication No. US20020161196A1  
GENERAL INFORMATION:



APPLICANT: Aihemri, Emad S.  
 TITLE OF INVENTION: NOVEL TRAIL RECEPTORS, NUCLEIC ACIDS ENCODING SAME, AND  
 FILE OF INVENTION: METHODS OF USE THEREOF  
 FILE REFERENCE: 480140.432D2  
 CURRENT APPLICATION NUMBER: US/10/076,773  
 CURRENT FILING DATE: 2002-02-12  
 NUMBER OF SEQ ID NOS: 6  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO: 6  
 LENGTH: 350  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-076-773-6

Query Match 45.7%; Score 48; DB 13; Length 350;  
 Best Local Similarity 43.8%; Pred. No. 55;  
 Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 PAKGMSPPGIVGEEG 17  
 Db 76 PSEGLCPGHHISEDG 91

RESULT 15  
 US-08-916-625B-2

Sequence 2; Application US/08916625B  
 Publication No. US20010010924A1  
 GENERAL INFORMATION:  
 APPLICANT: DEEN, KEITH C.  
 APPLICANT: YOUNG, PETER R.  
 TITLE OF INVENTION: TUMOR NECROSIS FACTOR RELATED  
 TITLE OF INVENTION: RECEPTOR, TR6  
 NUMBER OF SEQUENCES: 6  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: RATNER & PRESTIA  
 STREET: P.O. BOX 980  
 CITY: VALLEY FORGE  
 STATE: PA  
 COUNTRY: USA  
 ZIP: 19482  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/916,625B  
 FILING DATE: 22-AUG-1997  
 CLASSIFICATION: 536  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/853,684  
 FILING DATE: 09-MAY-1997  
 APPLICATION NUMBER: 60/041,230  
 FILING DATE: 14-MARCH-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: PRESTIA, PAUL F.  
 REGISTRATION NUMBER: 23,031  
 REFERENCE/DOCKET NUMBER: GH-50008-1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 610-407-0700  
 TELEFAX: 610-407-0701  
 TELEX: 846169  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 411 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-916-625B-2

Query Match 45.7%; Score 48; DB 8; Length 411;  
 Best Local Similarity 43.8%; Pred. No. 65;

Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 PAKGMSPPGIVGEEG 17  
 Db 76 PSEGLCPGHHISEDG 91

Search completed: March 22, 2004, 07:45:48  
 Job time: 4.09116 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:42:54 ; Search time 1.61201 Seconds  
(without alignments)  
640,518 Million cell updates/sec

Title: US-09-662-293-4

Sequence: 1 DPAKMSPPGFIVEEGVLS 20

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database:

1: Issued Patents AA:\*  
2: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/6C\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/backfill6B1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	100.0	20	4	US-09-292-225-4
2	105	100.0	20	4	US-09-292-225-13
3	101	96.2	430	4	US-09-292-225-41
4	101	96.2	509	4	US-09-292-225-35
5	101	96.2	509	4	US-09-292-225-38
6	99	94.3	536	4	US-09-292-225-21
7	99	94.3	555	4	US-09-292-225-15
8	99	94.3	555	4	US-09-292-225-18
9	98	45.7	350	4	US-09-134-618-6
10	98	45.7	411	3	US-09-329-633A-2
11	98	45.7	411	4	US-09-079-029-1
12	98	45.7	411	4	US-09-134-618-2
13	98	45.7	411	4	US-10-039-785-3
14	98	45.7	412	4	US-09-333-593A-2
15	98	45.7	424	4	US-09-333-593A-8
16	98	45.7	440	3	US-08-883-036A-2
17	98	45.7	440	4	US-09-536-201-2
18	98	45.7	440	4	US-09-578-392-2
19	98	45.7	440	4	US-09-252-991A-24263
20	98	45.7	440	4	US-09-252-991A-30682
21	98	45.7	440	4	US-09-252-991A-20817
22	98	45.7	440	4	US-08-705-660-34
23	98	45.7	440	4	US-08-989-045-34
24	98	45.7	440	4	US-09-252-991A-18602
25	98	45.7	440	4	US-09-543-681A-5334
26	98	45.7	440	4	US-09-252-991A-16628
27	98	45.7	440	4	US-09-134-000C-4952

28	41	39.0	433	4	US-09-252-991A-21851	Sequence 21851, A
29	41	39.0	453	4	US-09-711-164-349	Sequence 349, App
30	41	39.0	456	4	US-09-543-681A-5516	Sequence 5516, App
31	41	39.0	519	4	US-09-489-039A-11591	Sequence 11591, A
32	41	39.0	611	2	US-08-821-119-17	Sequence 119, A
33	41	39.0	611	2	US-08-821-118-5	Sequence 5, App1
34	41	39.0	845	4	US-09-198-452A-158	Sequence 458, App
35	41	39.0	1049	4	US-09-252-991A-17298	Sequence 17298, A
36	41	39.0	1596	4	US-09-328-352-5542	Sequence 5542, App
37	41	39.0	3165	2	US-08-459-146-3	Sequence 3, App1
38	41	39.0	3165	2	US-08-459-065-3	Sequence 3, App1
39	40.5	38.6	72	4	US-09-543-681A-5442	Sequence 5442, App
40	40.5	38.6	366	3	US-09-043-627-6	Sequence 6, App1
41	40.5	38.6	484	4	US-09-252-991A-32847	Sequence 32847, A
42	40.5	38.6	509	4	US-09-252-991A-26412	Sequence 26412, A
43	40	38.1	54	1	US-08-401-512-53	Sequence 53, App1
44	40	38.1	61	4	US-09-107-532A-6265	Sequence 6265, App
45	40	38.1	280	4	US-09-252-991A-21650	Sequence 21650, A

#### ALIGNMENTS

```
RESULT 1
US-09-292-225-4
Sequence 4, Application US/09292225
Patent No. 6455686
GENERAL INFORMATION:
APPLICANT: McCall, Catherine A.
APPLICANT: Hunter, Shirley Wu
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
FILE REFERENCE: AL-2-C3
CURRENT APPLICATION NUMBER: US/09/292,225
EARLIER FILING DATE: 1999-04-15
EARLIER APPLICATION NUMBER: 60/098,909
EARLIER FILING DATE: 1998-09-02
EARLIER APPLICATION NUMBER: 60/085,295
EARLIER FILING DATE: 1998-05-13
EARLIER APPLICATION NUMBER: 60/098,565
EARLIER FILING DATE: 1998-04-17
EARLIER APPLICATION NUMBER: 09/062,013
EARLIER FILING DATE: 1998-04-17
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 20
TYPE: PRT
ORGANISM: Dermatophagoides farinae
US-09-292-225-4

Query Match          100.0%; Score 105; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.6e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPAKMSPPGFIVEEGVLS 20
Db 1 DPAKMSPPGFIVEEGVLS 20

RESULT 2
US-09-292-225-13
Sequence 13, Application US/09292225
Patent No. 6455686
GENERAL INFORMATION:
APPLICANT: McCall, Catherine A.
APPLICANT: Hunter, Shirley Wu
APPLICANT: Weber, Eric R.
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
FILE REFERENCE: AL-2-C3
CURRENT APPLICATION NUMBER: US/09/292,225
```

CURRENT FILING DATE: 1999-04-15  
; EARLIER APPLICATION NUMBER: 60/098,909  
; EARLIER FILING DATE: 1998-09-02  
; EARLIER APPLICATION NUMBER: 60/085,295  
; EARLIER FILING DATE: 1998-05-13  
; EARLIER APPLICATION NUMBER: 60/098,565  
; EARLIER FILING DATE: 1998-04-17  
; EARLIER APPLICATION NUMBER: 09/062,013  
; EARLIER FILING DATE: 1998-04-17  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Dermatophagoides farinae  
US-09-292-225-13

Query Match 100.0%; Score 105; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.6e-10;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPAKMSPPGFIVGEGVLS 20  
|||||  
Db 1 DPAKMSPPGFIVGEGVLS 20

RESULT 3  
US-09-292-225-41  
; Sequence 41, Application US/09292225  
; Patent No. 6455686  
; GENERAL INFORMATION:  
; APPLICANT: McCall, Catherine A.  
; APPLICANT: Hunter, Shirley Wu  
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: AL-2-C3  
; CURRENT APPLICATION NUMBER: US/09/292,225  
; CURRENT FILING DATE: 1999-04-15  
; EARLIER APPLICATION NUMBER: 60/098,909  
; EARLIER FILING DATE: 1998-09-02  
; EARLIER APPLICATION NUMBER: 60/085,295  
; EARLIER FILING DATE: 1998-05-13  
; EARLIER APPLICATION NUMBER: 60/098,565  
; EARLIER FILING DATE: 1998-04-17  
; EARLIER APPLICATION NUMBER: 09/062,013  
; EARLIER FILING DATE: 1998-04-17  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 41  
; LENGTH: 490  
; TYPE: PRT  
; ORGANISM: Dermatophagoides farinae  
US-09-292-225-41

Query Match 96.2%; Score 101; DB 4; Length 490;  
Best Local Similarity 95.0%; Pred. No. 2.7e-08;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DPAKMSPPGFIVGEGVLS 20  
|||||  
Db 279 DPAKMSPPGFIVGEGVLS 298

RESULT 4  
US-09-292-225-35  
; Sequence 35, Application US/09292225  
; Patent No. 6455686  
; GENERAL INFORMATION:  
; APPLICANT: McCall, Catherine A.  
; APPLICANT: Hunter, Shirley Wu  
; APPLICANT: Weber, Eric R.  
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS

; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: AL-2-C3  
; CURRENT APPLICATION NUMBER: US/09/292,225  
; CURRENT FILING DATE: 1999-04-15  
; EARLIER APPLICATION NUMBER: 60/098,909  
; EARLIER FILING DATE: 1998-09-02  
; EARLIER APPLICATION NUMBER: 60/085,295  
; EARLIER FILING DATE: 1998-05-13  
; EARLIER APPLICATION NUMBER: 60/098,565  
; EARLIER FILING DATE: 1998-04-17  
; EARLIER APPLICATION NUMBER: 09/062,013  
; EARLIER FILING DATE: 1998-04-17  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 35  
; LENGTH: 509  
; TYPE: PRT  
; ORGANISM: Dermatophagoides farinae  
US-09-292-225-35

Query Match 96.2%; Score 101; DB 4; Length 509;  
Best Local Similarity 95.0%; Pred. No. 2.8e-08;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DPAKMSPPGFIVGEGVLS 20  
|||||  
Db 298 DPAKMSPPGFIVGEGVLS 317

RESULT 5  
US-09-292-225-38  
; Sequence 38, Application US/09292225  
; Patent No. 6455686  
; GENERAL INFORMATION:  
; APPLICANT: McCall, Catherine A.  
; APPLICANT: Hunter, Shirley Wu  
; APPLICANT: Weber, Eric R.  
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: AL-2-C3  
; CURRENT APPLICATION NUMBER: US/09/292,225  
; CURRENT FILING DATE: 1999-04-15  
; EARLIER APPLICATION NUMBER: 60/098,909  
; EARLIER FILING DATE: 1998-09-02  
; EARLIER APPLICATION NUMBER: 60/085,295  
; EARLIER FILING DATE: 1998-05-13  
; EARLIER APPLICATION NUMBER: 60/098,565  
; EARLIER FILING DATE: 1998-04-17  
; EARLIER APPLICATION NUMBER: 09/062,013  
; EARLIER FILING DATE: 1998-04-17  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 38  
; LENGTH: 509  
; TYPE: PRT  
; ORGANISM: Dermatophagoides farinae  
US-09-292-225-38

Query Match 96.2%; Score 101; DB 4; Length 509;  
Best Local Similarity 95.0%; Pred. No. 2.8e-08;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DPAKMSPPGFIVGEGVLS 20  
|||||  
Db 298 DPAKMSPPGFIVGEGVLS 317

RESULT 6  
US-09-292-225-21  
; Sequence 21, Application US/09292225  
; Patent No. 6455686  
; GENERAL INFORMATION:  
; APPLICANT: McCall, Catherine A.

```

; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; CURRENT FILING DATE: 1998-04-15
; EARLIER APPLICATION NUMBER: 60/098,909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085,295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098,565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062,013
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-09-292-225-21

Query Match          94.3%; Score 99; DB 4; Length 536;
Best Local Similarity 95.0%; Pred. No. 6.3e-08;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DPAKMSPPGFIVGEGVLS 20
DB 279 DPAKMSPPGFIVGEGVLS 298

RESULT 7
US-09-292-225-15
; Sequence 15, Application US/09292225
; Patent No. 6455686
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; CURRENT FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: 60/098,909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085,295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098,565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062,013
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-09-292-225-15

Query Match          94.3%; Score 99; DB 4; Length 555;
Best Local Similarity 95.0%; Pred. No. 6.6e-08;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DPAKMSPPGFIVGEGVLS 20
DB 298 DPAKMSPPGFIVGEGVLS 317
```

```

; Patent No. 6455686
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; CURRENT FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: 60/098,909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085,295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098,565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062,013
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-09-292-225-18

Query Match          94.3%; Score 99; DB 4; Length 555;
Best Local Similarity 95.0%; Pred. No. 6.6e-08;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DPAKMSPPGFIVGEGVLS 20
DB 298 DPAKMSPPGFIVGEGVLS 317

RESULT 9
US-09-134-618-6
; Sequence 6, Application US/09134618
; Patent No. 6417328
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Emad S.
; TITLE OF INVENTION: NOVEL TRAIL RECEPTORS, NUCLEIC ACIDS ENCODING SAME, AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 480140.432
; CURRENT APPLICATION NUMBER: US/09/134,618
; CURRENT FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-134-618-6

Query Match          45.7%; Score 48; DB 4; Length 350;
Best Local Similarity 43.8%; Pred. No. 6.7;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 PAKGMSPPGFIVGEGV 17
DB 76 PSEGLCPGHHISEDG 91

RESULT 10
US-09-329-633A-2
; Sequence 2, Application US/09329633A
; Patent No. 6252050
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Chantcharapai, Anan
; APPLICANT: Kim, K. Jin
; TITLE OF INVENTION: METHOD FOR MAKING MONOCLONAL ANTIBODIES AND
; TITLE OF INVENTION: CROSS-REACTIVE ANTIBODIES OBTAINABLE BY THE METHOD
```

FILE REFERENCE: P14681 (REVISED)  
CURRENT APPLICATION NUMBER: US/09/329,633A  
CURRENT FILING DATE: 1999-06-10  
PRIOR APPLICATION NUMBER: US 60/089,253  
PRIOR FILING DATE: 1998-06-12  
NUMBER OF SEQ ID NOS: 2  
SEQ ID NO 2  
LENGTH: 411  
TYPE: PRT  
ORGANISM: human  
FEATURE:  
NAME/KEY: xaa  
LOCATION: 410  
OTHER INFORMATION: xaa = leu or met  
US-09-329-633A-2

Query Match 45.7%; Score 48; DB 3; Length 411;  
Best Local Similarity 43.8%; Pred. No. 8;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 PAKGSPPGFIVEEG 17  
Db 76 PSEGLCPGHIHSDG 91

RESULT 11  
US-09-079-029-1  
Sequence 1, Application US/09079029  
Patent No. 6342369  
GENERAL INFORMATION:  
APPLICANT: Adams, Camilla W.  
APPLICANT: Ashkenazi, Avi J.  
APPLICANT: Chutcharapai, Anan  
APPLICANT: Kim, Kyung J.  
TITLE OF INVENTION: Apo-2 Receptor  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/079,029  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Marschang, Diane L.  
REGISTRATION NUMBER: 35,600  
REFERENCE/DOCKET NUMBER: P1101R2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-5416  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 411 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear  
US-09-079-029-1

Query Match 45.7%; Score 48; DB 4; Length 411;  
Best Local Similarity 43.8%; Pred. No. 8;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 PAKGSPPGFIVEEG 17  
Db 76 PSEGLCPGHIHSDG 91

RESULT 12  
US-09-134-618-2  
Sequence 2, Application US/09134618  
Patent No. 6417328  
GENERAL INFORMATION:  
APPLICANT: Alnemri, Emad S.  
TITLE OF INVENTION: NOVEL TRAIL RECEPTORS, NUCLEIC ACIDS ENCODING SAME, AND  
METHODS OF USE THEREOF  
FILE REFERENCE: 480140.432  
CURRENT APPLICATION NUMBER: US/09/134,618  
CURRENT FILING DATE: 1998-08-14  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 411  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-134-618-2

Query Match 45.7%; Score 48; DB 4; Length 411;  
Best Local Similarity 43.8%; Pred. No. 8;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 PAKGSPPGFIVEEG 17  
Db 76 PSEGLCPGHIHSDG 91

RESULT 13  
US-10-039-785-3  
Sequence 3, Application US/10039785  
Patent No. 6538938  
GENERAL INFORMATION:  
APPLICANT: Salcedo et al.  
TITLE OF INVENTION: Antibodies that Immunospecifically Bind to TRAIL  
FILE REFERENCE: PF550  
CURRENT APPLICATION NUMBER: US/10/039,785  
CURRENT FILING DATE: 2002-05-07  
PRIOR APPLICATION NUMBER: 60/369,860  
PRIOR FILING DATE: 2002-04-05  
PRIOR APPLICATION NUMBER: 60/341,237  
PRIOR FILING DATE: 2001-12-20  
PRIOR APPLICATION NUMBER: 60/331,310  
PRIOR FILING DATE: 2001-11-14  
PRIOR APPLICATION NUMBER: 60/331,044  
PRIOR FILING DATE: 2001-11-07  
PRIOR APPLICATION NUMBER: 60/327,364  
PRIOR FILING DATE: 2001-10-09  
PRIOR APPLICATION NUMBER: 60/323,807  
PRIOR FILING DATE: 2001-09-21  
PRIOR APPLICATION NUMBER: 60/309,176  
PRIOR FILING DATE: 2001-08-02  
PRIOR APPLICATION NUMBER: 60/294,981  
PRIOR FILING DATE: 2001-06-04  
PRIOR APPLICATION NUMBER: 60/293,473  
NUMBER OF SEQ ID NOS: 66  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3  
LENGTH: 411  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-039-785-3

Query Match 45.7%; Score 48; DB 4; Length 411;  
Best Local Similarity 43.8%; Pred. No. 8;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 PAKGSPPGFIVEEG 17  
Db 76 PSEGLCPGHIHSDG 91

Db 76 PSEGLCPGHHISEDG 91

RESULT 14  
US-09-333-593A-2  
Sequence 2, Application US/09333593A  
Patent No. 6313269  
GENERAL INFORMATION:  
APPLICANT: DEEN, KEITH C.  
APPLICANT: YOUNG, PETER R.  
APPLICANT: MARSHALL, LISA A.  
APPLICANT: ROSHAK, AMY K.  
APPLICANT: TAN, KONG B.  
APPLICANT: TRUNEH, ALEMSEGED  
TITLE OF INVENTION: TUMOR NECROSIS FACTOR RELATED RECEPTOR,  
TITLE OF INVENTION: TR6  
FILE REFERENCE: GH-50008-2  
CURRENT APPLICATION NUMBER: US/09/333,593A  
CURRENT FILING DATE: 1999-06-15  
PRIOR APPLICATION NUMBER: 08/916,625  
PRIOR FILING DATE: 1997-08-22  
PRIOR APPLICATION NUMBER: 08/853,684  
PRIOR FILING DATE: 1997-05-09  
PRIOR APPLICATION NUMBER: 60/041,230  
PRIOR FILING DATE: 1997-03-14  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 412  
TYPE: PRT  
ORGANISM: HOMO SAPIENS  
US-09-333-593A-2

Query Match 45.7%; Score 48; DB 4; Length 412;  
Best Local Similarity 43.8%; Pred. No. 8;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 PAKGMSPPGFIVGEEG 17  
Db 76 PSEGLCPGHHISEDG 91

RESULT 15  
US-09-333-593A-8  
Sequence 8, Application US/09333593A  
Patent No. 6313269  
GENERAL INFORMATION:  
APPLICANT: DEEN, KEITH C.  
APPLICANT: YOUNG, PETER R.  
APPLICANT: MARSHALL, LISA A.  
APPLICANT: ROSHAK, AMY K.  
APPLICANT: TAN, KONG B.  
APPLICANT: TRUNEH, ALEMSEGED  
TITLE OF INVENTION: TUMOR NECROSIS FACTOR RELATED RECEPTOR,  
TITLE OF INVENTION: TR6  
FILE REFERENCE: GH-50008-2  
CURRENT APPLICATION NUMBER: US/09/333,593A  
CURRENT FILING DATE: 1999-06-15  
PRIOR APPLICATION NUMBER: 08/916,625  
PRIOR FILING DATE: 1997-08-22  
PRIOR APPLICATION NUMBER: 08/853,684  
PRIOR FILING DATE: 1997-05-09  
PRIOR APPLICATION NUMBER: 60/041,230  
PRIOR FILING DATE: 1997-03-14  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 8  
LENGTH: 424  
TYPE: PRT  
ORGANISM: HOMO SAPIENS  
US-09-333-593A-8

Query Match 45.7%; Score 48; DB 4; Length 424;

Best Local Similarity 43.8%; Pred. No. 8.3;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;  
QY 2 PAKGMSPPGFIVGEEG 17  
Db 76 PSEGLCPGHHISEDG 91

Search completed: March 22, 2004, 07:03:54  
Job time : 2.61201 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:40:28 ; Search time 0.780434 seconds  
(without alignments)  
1479.047 Million cell updates/sec

Title: US-09-662-293-5  
Perfect score: 67  
Sequence: 1 DEKNSPECTLGP 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 segs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:  
1: p1r1:  
2: p1r2:  
3: p1r3:  
4: p1r4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	70.1	180	2 F88028	protein F46F5.15 (
2	42	62.7	181	2 T20324	hypothetical prote
3	41	61.2	291	2 F82938	iron(III) diclitrat
4	40	59.7	39	2 S70808	IVI protein VI - V
5	40	59.7	403	2 H82440	ABC transporter, A
6	39	58.2	187	2 G69164	hypothetical prote
7	39	58.2	213	2 A84995	crotonate phosphorib
8	39	58.2	234	2 AB2221	hypothetical prote
9	39	58.2	248	2 F72604	probable high-affi
10	39	58.2	2531	2 T31070	notch homolog - se
11	39	58.2	2703	1 A24420	notch protein - fr
12	38	56.7	141	2 G72100	hypothetical prote
13	38	56.7	141	2 C86521	hypothetical prote
14	38	56.7	306	2 T28388	ORF MSV227 leucine
15	38	56.7	334	2 T36475	probable DNA-bindi
16	38	56.7	380	2 T04343	L-ascorbate oxidas
17	38	56.7	1100	2 AB3243	conjugal transfer
18	38	56.7	1101	2 T03419	trial protein - Agr
19	38	56.7	1354	2 T13930	tripetidyl-peptid
20	37.5	56.0	166	2 S71209	ubiquitin-protein
21	37.5	55.2	166	2 T45703	ubiquitin-protein
22	37	55.2	30	2 A33308	chromomodulin - r
23	37	55.2	107	2 T00143	hypothetical 12.4k
24	37	55.2	164	2 D42148	GTP-binding protei
25	37	55.2	272	2 H97258	acetyl esterase fa
26	37	55.2	279	2 A96913	ABC-type sulfatase
27	37	55.2	307	1 G64717	pyridoxal phosphat
28	37	55.2	307	2 D71801	pyridoxal phosphat
29	37	55.2	414	2 F96553	unknown protein, 3

30	37	55.2	430	2 T18475	hypothetical prote
31	37	55.2	432	2 T18481	hypothetical prote
32	37	55.2	540	2 E75633	conserved hypothet
33	37	55.2	575	1 TH408	chromomodulin pre
34	37	55.2	577	2 A60501	chromomodulin pre
35	37	55.2	670	2 S47042	protein kinase (EC
36	37	55.2	686	2 JCT569	Delta-4 protein -
37	37	55.2	1021	2 T15765	hypothetical prote
38	37	55.2	1112	2 T47784	hypothetical prote
39	37	55.2	1492	2 T18560	DNA-directed DNA p
40	37	55.2	1513	2 T28158	probable DNA-direc
41	37	55.2	1597	2 S68420	citron - mouse
42	37	55.2	4180	2 G83559	hypothetical prote
43	37	53.7	97	2 AF0509	citrate lyase acyl
44	36	53.7	139	2 T03184	probable high-mobi
45	36	53.7	225	2 E69350	hypothetical prote

## ALIGNMENTS

RESULT 1  
F88028  
protein F46F5.15 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001  
C:Accession: F88028  
R:Anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolo  
A:Reference number: A75000; WUID:99069613; PMID:9851916  
A>Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/projects/C.e1  
A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; an  
A:Accession: F88028  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-180 <STO>  
A:Cross-references: GB:chr\_II; PIDN:ACT8194.1; FID:G3886043; GSPDB:GN00020; CESP:F46F5.  
A:Genetics:  
A:Gene: F46F5.15  
A:Map position: 2

Query Match  
Best Local Similarity 70.1%; Score 47; DB 2; Length 180;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 KNSPECTLGP 12  
DB 124 KGSFQICLGP 133

RESULT 2  
T20324  
hypothetical protein D1086.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T20324  
R:Stuy, R.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19257  
A:Accession: T20324  
A>Status: preliminary; translated from GB/EMBL/DBU  
A:Molecule type: DNA  
A:Residues: 1-181 <WII>  
A:Cross-references: EMBL:Z81491; PIDN:CAB04020.1; GSPDB:GN00023; CESP:D1086.2  
A:Experimental source: clone D1086  
C:Genetics:  
A:Gene: CESP:D1086.2  
A:Map position: 5  
A:Introns: 42/3; 135/3

Query Match  
Best Local Similarity 72.7%; Score 42; DB 2; Length 181;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 DEKNSFECTILG 11  
|||  
52 DEKNSFEMLRG 62

## RESULT 3

F82938

Iron(III) dicitrate transport ATP-binding protein U0069 [imported] - Ureaplasma urealyticum  
C:Species: Ureaplasma urealyticum

C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000

C/Accession: F82938

R/Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.  
submitted to Genbank, February 2000

A/Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mir

A/Reference number: A82870

A/Accession: F82938

A/Status: preliminary  
A/Molecule type: DNA

A/Residues: 1-291 <GIA>  
A/Cross-references: GB:AE002106; GB:AF222894; NID:G6899011; PIDN:AAF30474.1; GSPDB:GN001

C/Genetics:  
A/Experimental source: serovar 3; biovar 1

A/Gene: fecE; U0069  
A/Genetic code: SGC3

Query Match 61.2%; Score 41; DB 2; Length 291;

Best Local Similarity 72.7%; Pred. No. 7.6;

Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EKNSFECTILGP 12  
|||  
53 EKNSFTTILGP 63

## RESULT 4

S70808

Ivi protein VI - Vibrio cholerae (fragment)  
N/Alternate names: cysA protein homolog

C/Species: Vibrio cholerae

C>Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 08-Oct-1999

C/Accession: S70808

R/Camilli, A.; Mektanos, J.J.  
Mol. Microbiol. 18, 671-683, 1995

A/Title: Use of recombinase gene fusions to identify Vibrio cholerae genes induced durin

A/Reference number: S70798; MUID:96414469; PMID:8817490

A/Accession: S70808

A/Status: preliminary; nucleic acid sequence not shown  
A/Molecule type: DNA

A/Residues: 1-39 <CAN>

A/Cross-references: EMBL:U25729; NID:G1165187; PIDN:AA43557.1; PID:G1165188

Query Match 59.7%; Score 40; DB 2; Length 39;  
Best Local Similarity 63.6%; Pred. No. 1.5;

Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 EKNSFECTILGP 12  
|||  
6 EKGEFVCFILGP 16

## RESULT 5

H82440

ABC transporter, ATP-binding protein VCA0602 [imported] - Vibrio cholerae (strain N16961  
C:Species: Vibrio cholerae

C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 03-Jun-2002

C/Accession: H82440  
R/Heideberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gilm, M.L.; Dodson, R.U.;  
Hardison, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, E.

1, R.R.; Mektanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000

A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A/Reference number: A82035; MUID:20406833; PMID:10952301

A/Accession: H82440

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-403 <HEI>

A/Cross-references: GB:AE004390; GB:AE003853; NID:9658003; PIDN:AAF96503.1; GSPDB:GN001

A/Experimental source: serogroup O1, strain N16961, biotype El Tor

C/Genetics:

A/Gene: VCA0602

C/Map position: 2

C/Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 59.7%; Score 40; DB 2; Length 403;  
Best Local Similarity 63.6%; Pred. No. 1.6;

Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 EKNSFECTILGP 12  
|||  
58 EKGEFVCFILGP 68

## RESULT 6

G69164

hypothetical protein MTH493 - Methanobacterium thermoautotrophicum (strain Delta H)  
C/Species: Methanobacterium thermoautotrophicum

C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 04-Mar-2000

C/Accession: G69164

R/Smith, D.R.; Doucette-Stamm, L.A.; Delouphery, C.; Lee, H.; Dubois, J.; Aldredge, T.;  
Qin, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.

1, S.; Church, G.M.; Daniels, C.D.; Mao, J.; Rice, P.; Noelling, D.; Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997

A/Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct

A/Reference number: A69000; MUID:98037514; PMID:9371463

A/Accession: G69164

A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA

A/Residues: 1-187 <MTH>

A/Cross-references: GB:AE000833; GB:AE000666; NID:G2621559; PIDN:AB84999.1; PID:G262156

A/Experimental source: strain Delta H

C/Genetics:

A/Gene: MTH493

A/Start codon: TTG

C/Superfamily: Methanobacterium thermoautotrophicum hypothetical protein MTH493

Query Match 58.2%; Score 39; DB 2; Length 187;  
Best Local Similarity 54.5%; Pred. No. 1.1;

Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 DEKNSFECTILG 11  
|||  
28 DEHNMICVLG 38

## RESULT 7

A84995

orotate phosphoribosyltransferase (EC 2.4.2.10) [imported] - Buchnera sp. (strain APS)  
C/Species: Buchnera sp.

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001

C/Accession: A84995  
R/Shigenobu, S.; Mitanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.

Nature 407, 81-86, 2000

A/Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. A

A/Reference number: A84930; MUID:2045173; PMID:10993077

A/Accession: A84995

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-213 <STO>  
A/Cross-references: GB:AP000398; GSPDB:GN00144

A/Experimental source: strain APS  
C/Genetics:  
A/Gene: PYR2; BU559

C/Superfamily: orotate phosphoribosyltransferase; orotate phosphoribosyltransferase homo  
C/Keywords: glycosyltransferase; pentosyltransferase



```

Query Match      58.2%; Score 39; DB 2; Length 213;
Best Local Similarity 50.0%; Pred. No. 13;
Matches      6; Conservative      2; Mismatches      4; Indels      0; Gaps      0;

QY      1 DEKNSPECILGP 12
      ||| ||| |||
Db      58 DSKKRFVLPFGP 69

RESULT 8
AB2221
hypothetical protein a113321 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AB2221
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kunitz, T.; Sasamoto, S.; Matsumoto, A.; Iritani, S.;
Nakazaki, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AB2221
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-234 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA075020.1; PID:G17132416; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: a113321

Query Match      58.2%; Score 39; DB 2; Length 234;
Best Local Similarity 66.7%; Pred. No. 14;
Matches      6; Conservative      2; Mismatches      1; Indels      0; Gaps      0;

QY      2 EKNSPECIL 10
      ||| ||| |||
Db      124 EKNGFDCL 132

RESULT 9
F72604
probable high-affinity branched-chain amino acid transport ATP-binding protein APE1301 -
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 02-Sep-2000
C:Accession: F72604
R:Kawabuchi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A:Reference number: A72450; MUID:99310339; PMID:10382366
A:Accession: F72604
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-248 <KAM>
A:Cross-references: DDBJ:AF000061; NID:G5104821; PIDN:BA080292.1; PID:d1044078; PID:G510
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1301
C:Superfamily: short-chain ATP-binding cassette proteins; ATP-binding cassette homology
F:9-205/Domain: ATP-binding cassette homology <ABC>

Query Match      58.2%; Score 39; DB 2; Length 248;
Best Local Similarity 54.5%; Pred. No. 15;
Matches      6; Conservative      2; Mismatches      3; Indels      0; Gaps      0;

QY      2 EKNSPECILGP 12
      ||| ||| |||
Db      17 EKGSITCLIP 27

RESULT 10
T31070
notch homolog - sea urchin (Lytechinus variegatus)
C:Species: Lytechinus variegatus (variegated urchin)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 31-Jan-2000
C:Accession: T31070
R:Sherwood, D.R.; McClay, D.R.
Development 124, 3363-3374, 1997
A:Title: Identification and localization of a sea urchin Notch homologue: insights into
A:Reference number: Z20966; MUID:97454256; PMID:9310331
A:Accession: T31070
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2531 <SHR>
A:Cross-references: EMBL:AF000634; NID:G2570350; PID:G2570351; PIDN:AA082088.1
C:Superfamily: notch protein; ankyrin repeat homology; EGF homology

Query Match      58.2%; Score 39; DB 2; Length 2531;
Best Local Similarity 66.7%; Pred. No. 16e+02;
Matches      6; Conservative      2; Mismatches      1; Indels      0; Gaps      0;

QY      1 DEKNSPECT 9
      ||| ||| |||
Db      796 DEVNSFOCV 804

RESULT 11
A24420
notch protein - fruit fly (Drosophila melanogaster)
N:Alternate names: neurogenic repetitive locus protein
C:Species: Drosophila melanogaster
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A24420; A24768; S09358; A05267
R:Kidd, S.; Kelley, M.R.; Young, M.W.
Mol. Cell. Biol. 6, 3094-3108, 1986
A:Reference number: A24420; MUID:87064624; PMID:3097517
A:Accession: A24420
A:Molecule type: DNA
A:Residues: 1-2703 <KID>
A:Cross-references: GB:K03508; NID:G157991; PIDN:AA08725.1; PID:G157993
R:Wharton, K.A.; Johansen, K.M.; Xu, T.; Artavanis-Tsakonas, S.
Cell 43, 567-581, 1985
A:Reference number: A24768; MUID:86079539; PMID:3935325
A:Accession: A24768
A:Molecule type: mRNA
A:Residues: 1-48 'I', 50-118, 'R', 120-230, 'I', 232-256, 'N', 258-266, 'A', 268-872, 'R', 874-958
A:Note: the authors translated the codon ATC for residue 49 as Thr, ATT for residue 204
R:Itauz, D.
Nucleic Acids Res. 17, 6463-6471, 1989
A:Title: Hypervariability of simple sequences as a general source for polymorphic DNA m
A:Reference number: S09358; MUID:89385974; PMID:2780284
A:Accession: S09358
A:Molecule type: DNA
A:Residues: 2505-2551, 'QQQQ', 2552-2576, 'E', 2578-2604 <TAU>
R:Wharton, K.A.; Yedvobnick, B.; Finnerty, V.G.; Artavanis-Tsakonas, S.
Cell 40, 55-62, 1985
A:Title: opa: a novel family of transcribed repeats shared by the Notch locus and other
A:Reference number: A05267; MUID:85099329; PMID:2981631
A:Accession: A05267
A:Molecule type: DNA
A:Residues: 2504-2576, 'E', 2578-2611 <WHA2>
C:Genetics:
A:Gene: notch, opa
A:Cross-references: FlyBase:FBgn0004647
A:Map position: 8, 96-9.36
A:Intron: 53/3; 84/3; 171/3; 240/3; 283/3; 2333/3; 2436/3; 2588/3
C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
C:Keywords: differentiation; tandem repeat; transmembrane protein
F:27-43/Domain: transmembrane #status predicted <TM1>
F:297-328/Domain: EGF homology <EGF1>
F:530-561/Domain: EGF homology <EGF1>
F:568-599/Domain: EGF homology <EGF>
F:988-1019/Domain: EGF homology <EGX2>
F:1064-1095/Domain: EGF homology <EGF3>
F:1187-1218/Domain: EGF homology <EGX3>

```

F:1746-1762/Domain: transmembrane #status predicted <TMW2>  
F:1950-1982/Domain: ankyrin repeat homology <AN1>  
F:1983-2015/Domain: ankyrin repeat homology <AN2>  
F:1988-2004/Domain: transmembrane #status predicted <TMW3>  
F:2017-2049/Domain: ankyrin repeat homology <AN3>  
F:2050-2082/Domain: ankyrin repeat homology <AN4>  
F:2083-2115/Domain: ankyrin repeat homology <AN5>  
F:2558-2568/Region: glutamine-rich  
F:2558-2568/Domain: neurogenic repetitive element #status predicted <OPA>

Query Match 58.2%; Score 39; DB 1; Length 2703;  
Best Local Similarity 50.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 KNSFECTILGP 12  
DB 618 DVNSFKCLCDP 629

RESULT 12  
G72100  
hypothetical protein CP0519 [imported] - Chlamydomonas reinhardtii (strain CC-1690) and AF  
C:Species: Chlamydomonas reinhardtii, Chlamydomonas reinhardtii  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
C:Accession: G72100; H81567  
R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydomonas reinhardtii and C. trachomatis.  
A:Reference number: A72000; MUID:99206606; PMID:10192388  
A:Accession: G72100  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-141 <ARN>  
A:Cross-references: GB:AE001610; GB:AE001363; NID:94376515; PID:9437651  
A:Experimental source: strain CHL029  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, J.; McElarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydomonas reinhardtii and Chlamydomonas trachomatis.  
A:Reference number: A81500; MUID:20150255; PMID:10684935  
A:Accession: H81567  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-141 <RNA>  
A:Cross-references: GB:AE002212; GB:AE002161; NID:97189430; PID:AA93845.1; PID:G718943  
A:Experimental source: strain AR39, HL cells  
C:Genetics:  
A:Gene: CP0243; CP0519

Query Match 56.7%; Score 38; DB 2; Length 141;  
Best Local Similarity 50.0%; Pred. No. 13;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 KNSFECTILGP 12  
DB 76 KENFQCLFGP 85

RESULT 13  
G86521  
hypothetical protein CP0243 [imported] - Chlamydomonas reinhardtii (strain J138)  
C:Species: Chlamydomonas reinhardtii, Chlamydomonas reinhardtii  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
C:Accession: G86521  
R:Shira, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ie  
Nucleic Acids Res. 28, 2311-2314, 2000  
A:Title: Comparison of whole genome sequences of Chlamydomonas reinhardtii J138.  
A:Reference number: A86491; MUID:20330349; PMID:10871362  
A:Accession: G86521  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-141 <STO>  
A:Cross-references: GB:BA000068; NID:98978616; PID:BA98453.1; GSPDB:GN00142

A:Experimental source: strain J138  
C:Genetics:  
A:Gene: CP0243

Query Match 56.7%; Score 38; DB 2; Length 141;  
Best Local Similarity 50.0%; Pred. No. 13;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 KNSFECTILGP 12  
DB 76 KENFQCLFGP 85

RESULT 14  
T28388  
ORF MSV227 leucine rich repeat gene family protein - Melanoplus sanguinipes entomopoxvir  
C:Species: Melanoplus sanguinipes entomopoxvirus  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
C:Accession: T28388  
R:Alonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.  
J. Virol. 73, 533-552, 1999  
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.  
A:Reference number: Z20484; MUID:99102612; PMID:9847359  
A:Accession: T28388  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-306 <APO>  
A:Cross-references: EMBL:AF063866; NID:94049647; PID:AA97871.1; PID:94049911  
A:Genetics:  
A:Note: MSV227

Query Match 56.7%; Score 38; DB 2; Length 306;  
Best Local Similarity 77.8%; Pred. No. 28;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 KNSFECTILGP 10  
DB 8 EINSFECIM 16

RESULT 15  
T36475  
probable DNA-binding regulator protein - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T36475  
R:Seeger, K.D.; Harris, D.; Thomson, N.R.; Parkhill, J.; Barrell, B.G.; Randeram, M.A.  
submitted to the EMBL Data Library, September 1999  
A:Reference number: Z21607  
A:Accession: T36475  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-334 <SEB>  
A:Cross-references: EMBL:AL10470; PID:CA854172.1; GSPDB:GN00070; SCOPDB:SCF85.15  
C:Genetics:  
A:Gene: SCOPDB:SCF85.15

Query Match 56.7%; Score 38; DB 2; Length 334;  
Best Local Similarity 60.0%; Pred. No. 31;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 KNSFECTILGP 12  
DB 94 QNGRCVCLGP 103

Search completed: March 22, 2004, 07:01:23  
Job time: 2.78043 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:31:13 ; Search time 0.460256 Seconds

(without alignments) 1357.597 Million cell updates/sec

Title: US-09-662-293-5

Perfect score: 67

Sequence: 1 DEKNSPECTLGP 12

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	58.2	213	1	PYRE_BUCAL
2	39	58.2	2703	1	NOTC_DROME
3	38	56.7	1100	1	TRA_AGRIS
4	38	56.7	1242	1	JAGI_BRARE
5	37.5	56.0	166	1	UBCD_ARATH
6	37.5	56.0	166	1	UBCD_ARATH
7	37	55.2	164	1	RB13_RAT
8	37	55.2	307	1	PDXA_HELPJ
9	37	55.2	307	1	PDXA_HELPJ
10	37	55.2	575	1	TRBM_HUMAN
11	37	55.2	577	1	TRBM_MOUSE
12	37	55.2	664	1	PD13_MOUSE
13	37	55.2	686	1	DLA_MOUSE
14	37	55.2	1024	1	YL88_CABEL
15	37	55.2	1286	1	CTRO_HUMAN
16	37	55.2	1492	1	DPOA_OXYNO
17	37	55.2	1513	1	DPOA_OXYNO
18	37	55.2	1513	1	CTRO_MOUSE
19	37	55.2	1513	1	CTRO_MOUSE
20	36	53.7	252	1	ORC6_HUMAN
21	36	53.7	252	1	ORC6_MOUSE
22	36	53.7	346	1	DDL_CAMJE
23	36	53.7	385	1	KAPR_NEUCR
24	36	53.7	405	1	ENTW_BACFR
25	36	53.7	482	1	MTAI_ANASP
26	36	53.7	570	1	FBRP_STRPU
27	36	53.7	617	1	S6A2_HUMAN
28	36	53.7	664	1	PD13_RAT
29	36	53.7	1188	1	RPR2_ARATH
30	36	53.7	1191	1	RPR2_LYCES
31	36	53.7	1218	1	JAGI_MOUSE
32	35	52.2	89	1	RS17_LEPIN
33	35	52.2	97	1	CTLG_KLEPN

34	35	52.2	192	1	SCPI_PONLE	P05946 pontastacus
35	35	52.2	192	1	SCPI_PENSP	P02636 penaeus sp.
36	35	52.2	192	1	SCPI_PENSP	P02635 penaeus sp.
37	35	52.2	218	1	SC22_ARATH	Q94442 arabidopsis
38	35	52.2	254	1	YGC1_BACSU	P45944 bacillus su
39	35	52.2	263	1	YGC1_MERVA	O69045 methanococc
40	35	52.2	376	1	KAPR_CABEL	P30625 caenorhabdi
41	35	52.2	401	1	KAPR_CABEL	P30625 caenorhabdi
42	35	52.2	415	1	KAPR_CABEL	P30625 caenorhabdi
43	35	52.2	415	1	KAPR_CABEL	P30625 caenorhabdi
44	35	52.2	428	1	KAPR_CABEL	P30625 caenorhabdi
45	35	52.2	465	1	KAPR_CABEL	P30625 caenorhabdi

## ALIGNMENTS

Query Match	Score	DB 1	Length	213
Best Local Similarity	50.0%	Pred. No. 4.4		
Matches	6	Conservative	2	Mismatches
			4	Indels
			0	Gaps
			0	

QY 1 DEKONSECTINGP 12  
 Db 58 DSKXKFDVLPFGP 69

RESULT 2  
 NOTC DROME STANDARD; PRT; 2703 AA.  
 ID NOTC DROME STANDARD; PRT; 2703 AA.  
 AC P07267; 097458; P04154; 09W4tr8;  
 DT 01-NOV-1986 (Rel. 03, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Neurogenic locus Notch protein precursor.  
 GN N OR EG:140611.1 OR EG:163A10.2 OR CG3936.  
 OS Drosophila melanogaster (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OC NCBI\_TaxID=7227;  
 OX [1]  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Oregon-R; TISSUE=Embryo;  
 RX MEDLINE=86079539; PubMed=3935325;  
 RA Wharton K.A., Johansen K.M., Xu T., Artavanis-Tsakonas S.;  
 RT "Nucleotide sequence from the neurogenic locus notch implies a gene  
 product that shares homology with proteins containing EGF-like  
 repeats.";  
 RT Cell 43:567-581(1985).  
 RL [2]  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Canton-S. and Oregon-R; TISSUE=Embryo;  
 RX MEDLINE=87064624; PubMed=3097517;  
 RA Kidd S., Kelley M.R., Young M.W.;  
 RT "Sequence of the notch locus of Drosophila melanogaster: relationship  
 of the encoded protein to mammalian clotting and growth factors.";  
 RT Mol. Cell. Biol. 6:3094-3108(1986).  
 RL [3]  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Baxu A., Bakendale B., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.A., Bouck J., Brokstein P., Brotler P.,  
 RA Burttis K.C., Busam D.R., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hooten D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,  
 RA Jaitai M., Kalish F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kamel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lesko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Matei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskens D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Schaefer F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng U.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RT Science 287:2185-2195(2000).  
 RL [4]  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Oregon-R;  
 RX MEDLINE=20196011; PubMed=10731137;  
 RA Benos P.V., Galt M.K., Ashburner M., Murphy L., Harris D.,  
 RA Barrell B.G., Ferraz C., Vidal S., Brun C., Dematilles J., Cadieu E.,  
 RA Dreano S., Gloux S., Lelaure V., Mottier S., Gallibert F., Borokova D.,  
 RA Mirana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,  
 RA Paspigianakis G., Spanos L., Cox S., Madueno E., de Pablo B.,  
 RA Modellell J., Peter A., Schoettler P., Werner M., Mounkicli F.,  
 RA Belhert N., Dowe G., Schaefer U., Ueckle H., Bucheton A.,  
 RA Callister D.M., Campbell L.A., Darlmitou A., Henderson N.S.,  
 RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,  
 RA Glover D.M.;  
 RT "From sequence to chromosome: the tip of the X chromosome of D.  
 melanogaster.";  
 RT Science 287:2220-2222(2000).  
 RL [5]  
 RN [5]  
 RP SEQUENCE OF 2505-2611 FROM N.A.  
 RX MEDLINE=85099329; PubMed=2981631;  
 RA Wharton K.A., Yelovnick B., Finerty V.G., Artavanis-Tsakonas S.;  
 RT "opa: a novel family of transcribed repeats shared by the Notch locus  
 and other developmentally regulated loci in D. melanogaster.";  
 RT Cell 40:55-62(1985).  
 RL [6]  
 RN [6]  
 RP SEQUENCE OF 1-8 FROM N.A.  
 RX MEDLINE=87257846; PubMed=3037327;  
 RA Kelley M.R., Kidd S., Berg R.L., Young M.W.;  
 RT "Restriction of P-element insertions at the Notch locus of Drosophila  
 melanogaster.";  
 RT Mol. Cell. Biol. 7:1545-1548(1987).  
 RL [7]  
 RN [7]  
 RP INTERACTION WITH DX, AND MUTANT SU42C.  
 RX MEDLINE=94215489; PubMed=8162848;  
 RA Diederich R.J., Matsumo K., Ring H., Artavanis-Tsakonas S.;  
 RT "Cytosolic interaction between deltex and Notch ankyrin repeats  
 implicates deltex in the Notch signaling pathway.";  
 RT Development 120:473-481(1994).  
 RL [8]  
 RN [8]  
 RP INTERACTION WITH DX.  
 RX MEDLINE=95401878; PubMed=7671825;  
 RA Matsumo K., Diederich R.J., Go M.J., Blumweller C.M.,  
 RA Artavanis-Tsakonas S.;  
 RT "Deltex acts as a positive regulator of Notch signaling through  
 interactions with the Notch ankyrin repeats.";  
 RT Development 121:2633-2644(1995).  
 RL [9]  
 RN [9]  
 RP S3 CLEAVAGE BY PSN.  
 RX MEDLINE=99221487; PubMed=10206646;  
 RA Struhl G., Greenwald I.;  
 RT "Presenilin is required for activity and nuclear access of Notch in  
 Drosophila.";  
 RT Nature 398:522-525(1999).  
 RL [10]  
 RN [10]  
 RP S3 CLEAVAGE BY PSN.  
 RX MEDLINE=99221488; PubMed=10206647;  
 RA Ye Y., Luknova N., Fortini M.E.;  
 RT "Neurogenic phenotypes and altered Notch processing in Drosophila  
 Presenilin mutants.";  
 RT Nature 398:525-529(1999).  
 RL [11]  
 RN [11]  
 RP S2 CLEAVAGE BY KUZ.  
 RX MEDLINE=21657146; PubMed=11799064;  
 RA Lieber T., Kidd S., Young M.W.;  
 RT "Kuzbanian-mediated cleavage of Drosophila Notch.";  
 RT Genes Dev. 16:209-221(2002).  
 RL [12]

RP MUTANT MCD5.  
 RX MEDLINE=21575956; PubMed=11719214;  
 RA Raman P., Khechumian K., Seugnet U., Ardogaast N., Ackermann C.,  
 RA Heitzler P.;  
 RT "Novel Notch alleles reveal a Deltex-dependent pathway repressing  
 RT neural fate.";  
 RL Curr. Biol. 11:1729-1738(2001).  
 RN [13]  
 RP REVIEW.  
 RX MEDLINE=22256570; PubMed=12369105;  
 RA Portin P.;  
 RT "General outlines of the molecular genetics of the Notch signalling  
 RT pathway in *Drosophila melanogaster*: a review.";  
 RL Hereditas 136:89-96(2002).  
 CC -1- FUNCTION: Signaling protein, which regulates, with both positive  
 CC and negative signals, the differentiation of at least central and  
 CC peripheral nervous system and eye, wing disk, oogenesis, segmental  
 CC appendages such as antennae and legs, and muscles, through lateral  
 CC inhibition or induction. Functions as a receptor for membrane-  
 CC bound ligands Delta and Serrate to regulate cell-fate  
 CC determination. Upon ligand activation, and releasing from the cell  
 CC membrane, the Notch intracellular domain (NICD) forms a  
 CC transcriptional activator complex with Su(H) (suppressor of  
 CC haites) and activates genes of the E(Sp1) complex. Essential for  
 CC proper differentiation of ectoderm.  
 CC -1- SUBUNIT: Interacts with Su(H) when activated. Interacts with Dk  
 CC via its ANK repeats.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Upon activation and  
 CC 83 cleavage, it is released from the cell membrane and enters into  
 CC the nucleus in conjunction with Su(H).  
 CC -1- PTM: Upon binding its ligands such as Delta or Serrate, it is  
 CC cleaved (S2 cleavage) in its extracellular domain, close to the  
 CC transmembrane domain. S2 cleavage is probably mediated by Kex. It  
 CC is then cleaved (S3 cleavage) downstream of its transmembrane  
 CC domain, releasing it from the cell membrane. S3 cleavage requires  
 CC Psn.  
 CC -1- SIMILARITY: Belongs to the NOTCH family.  
 CC -1- SIMILARITY: Contains 36 EGF-like domains.  
 CC -1- SIMILARITY: Contains 3 Lin/Notch repeats.  
 CC -1- SIMILARITY: Contains 6 ANK repeats.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, M16152; AAB59220.1; -  
 DR EMBL, M16153; AAB59220.1; JOINED.  
 DR EMBL, M16149; AAB59220.1; JOINED.  
 DR EMBL, M16150; AAB59220.1; JOINED.  
 DR EMBL, M16151; AAB59220.1; JOINED.  
 DR EMBL, K03508; AAA28725.1; -  
 DR EMBL, M13689; AAA28725.1; JOINED.  
 DR EMBL, K03507; AAA28725.1; JOINED.  
 DR EMBL, AE003426; AAF45848.2; -  
 DR EMBL, AL035436; CAB37610.1; -  
 DR EMBL, AL035395; CAB37610.1; JOINED.  
 DR EMBL, M12175; AAA74496.1; -  
 DR EMBL, M16025; AAA28726.1; -

Query Match 58.2%; Score 39; DB 1; Length 2703;  
 Best Local Similarity 50.0%; Pred. No. 61;  
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DEKNSPECTLGP 12  
 DB 618 DVVNSFKLCIDP 629

RESULT 3

TRAA\_AGR5  
 ID TRAA\_AGR5 STANDARD; PRT; 1100 AA.  
 AC Q4349;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Conjugal transfer protein traa.  
 GN TRAA OR ATU6127 OR AGR\_PTI\_237.  
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
 OG Plasmid pTIC58.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 CC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.  
 OX NCBI\_TaxID=176299;  
 RX MEDLINE=96312367; PubMed=8763953;  
 RA Farrand S.K., Hwang I., Cook D.M.;  
 RT "The tra region of the nopaline-type T1 plasmid is a chimera with  
 RT elements related to the transfer systems of RSF1010, RP4, and F.";  
 RL J. Bacteriol. 178:4233-4247(1996).  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21608550; PubMed=11743193;  
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,  
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boye D., Sr.,  
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,  
 RA Kutayir T., Levy R., Li M.-T., McClelland E., Palmeri A.,  
 RA Raymond C., Rouse G., Saenphummachak C., Wu Z., Romero P., Gordon D.,  
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.F., Olson M.V.,  
 RA Nester E.W.;  
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens  
 RT C58.";  
 RL Science 294:2317-2323(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21608551; PubMed=11743194;  
 RA Goodner B., Hinkle G., Gatlung S., Miller N., Blanchard M.,  
 RA Quicilo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,  
 RA Homail K., Gordon J., Vaudin M., Tarchook O., Bpp A., Liu F.,  
 RA Wolian C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,  
 RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,  
 RA Cielo C., Slater S.;  
 RT "Genome sequence of the plant pathogen and biotechnology agent  
 RT Agrobacterium tumefaciens C58.";  
 RL Science 294:2323-2328(2001).  
 CC -1- SIMILARITY: Belongs to the mobA/mobL family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, AF010180; AAC17212.1; -  
 DR EMBL, AE009431; AAL46363.1; -  
 DR EMBL, AE007939; AAK91091.1; ALT\_INIT.  
 DR PIR, AE3243; AE3243.  
 DR PIR, T03419; T03419.  
 DR InterPro: IPR005053; MobA\_MobL.  
 DR Pfam: PF03389; MobA\_MobL.  
 KM Conjugation; ATP-binding; Plasmid; Complete proteome.  
 FT NP\_BIND 404 411 ATP (POTENTIAL).  
 FT CONFLICT 372 372 L -> LL (IN REF. 1).  
 SQ SEQUENCE 1100 AA; 123477 MW; B85D31DB526B7344 CRC64;

Query Match 56.7%; Score 38; DB 1; Length 1100;  
 Best Local Similarity 50.0%; Pred. No. 38;  
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DENOSFECTIGP 12  
 Db 385 EOKAIEFCLNGP 396  
 RESULT 4  
 JAG1\_BRARE STANDARD; PRT; 1242 AA.  
 ID JAG1\_BRARE  
 AC Q90157;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DE JAGged 1 precursor (Jagged1).  
 GN JAG1.  
 OS Brachydanio rerio (Zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Oda T., Chandrasekharappa S.C.;  
 RT "Isolation, characterization and expression analysis of Zebrafish  
 RT Jaggged genes";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Ligand for multiple Notch receptors and involved in the  
 CC mediation of Notch signaling (By similarity). Seems to be involved  
 CC in cell-face decisions.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).  
 CC -1- SIMILARITY: Contains 17 EGF-like domains.  
 CC -1- SIMILARITY: Contains 1 DSL domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF229446; AAI08213.1; -  
 CC ZFTN; ZDB-GENE-011128-2; jag1.  
 DR InterPro: IPR00152; Aex\_Hydroxyl\_S.  
 DR InterPro: IPR00174; DSL.  
 DR InterPro: IPR000742; EGF\_2.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR001438; EGF\_T1.  
 DR InterPro: IPR006209; EGF\_T1.  
 DR InterPro: IPR002049; Laminin\_EGF.  
 DR InterPro: IPR009041; PMP\_inhibitor.  
 DR InterPro: IPR001007; VWF\_C.  
 DR Pfam; PF01414; DSL; 1.  
 DR Pfam; PF00008; EGF; 14.  
 DR PRINTS; PRO0010; EGFBLDOD.  
 DR PRINTS; PRO0011; EGFAMININ.  
 DR SMART; SMO0051; DSL; 1.  
 DR SMART; SMO0179; EGF\_Ca; 9.  
 DR SMART; SMO0214; VMC\_1.  
 DR PROSITE; PS00010; Aex\_Hydroxyl; 9.  
 DR PROSITE; PS00022; EGF\_1; 16.  
 DR PROSITE; PS01186; EGF\_2; 12.  
 DR PROSITE; PS00026; EGF\_3; 15.  
 DR PROSITE; PS01187; EGF\_Ca; 8.  
 KM Calcium-binding; EGF-Like domain; Glycoprotein; Developmental protein;  
 KW Repeat; Transmembrane; Signal.  
 FT SIGNAL 1 28  
 FT CHAIN 29 1242 JAGGED 1.  
 FT DOMAIN 29 1070 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1071 1095 POTENTIAL.  
 FT DOMAIN 1096 1242 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 168 230 DSL.  
 FT DOMAIN 231 264 EGF-Like 1.

FT DOMAIN 262 295 EGF-Like 2.  
 FT DOMAIN 297 335 EGF-Like 3.  
 FT DOMAIN 337 373 EGF-Like 4.  
 FT DOMAIN 375 411 EGF-Like 5, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 413 449 EGF-Like 6, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 451 486 EGF-Like 7, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 488 524 EGF-Like 8, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 526 562 EGF-Like 9.  
 FT DOMAIN 575 630 EGF-Like 10.  
 FT DOMAIN 632 668 EGF-Like 11, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 670 706 EGF-Like 12, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 708 744 EGF-Like 13.  
 FT DOMAIN 747 783 EGF-Like 14.  
 FT DOMAIN 785 821 EGF-Like 15, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 823 859 EGF-Like 16, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 917 959 EGF-Like 17.  
 FT DISULFID 235 246 BY SIMILARITY.  
 FT DISULFID 239 252 BY SIMILARITY.  
 FT DISULFID 254 263 BY SIMILARITY.  
 FT DISULFID 266 277 BY SIMILARITY.  
 FT DISULFID 272 283 BY SIMILARITY.  
 FT DISULFID 285 294 BY SIMILARITY.  
 FT DISULFID 301 313 BY SIMILARITY.  
 FT DISULFID 307 323 BY SIMILARITY.  
 FT DISULFID 325 334 BY SIMILARITY.  
 FT DISULFID 341 352 BY SIMILARITY.  
 FT DISULFID 346 361 BY SIMILARITY.  
 FT DISULFID 363 372 BY SIMILARITY.  
 FT DISULFID 379 390 BY SIMILARITY.  
 FT DISULFID 384 399 BY SIMILARITY.  
 FT DISULFID 401 410 BY SIMILARITY.  
 FT DISULFID 417 428 BY SIMILARITY.  
 FT DISULFID 422 437 BY SIMILARITY.  
 FT DISULFID 439 448 BY SIMILARITY.  
 FT DISULFID 455 465 BY SIMILARITY.  
 FT DISULFID 459 474 BY SIMILARITY.  
 FT DISULFID 476 485 BY SIMILARITY.  
 FT DISULFID 492 503 BY SIMILARITY.  
 FT DISULFID 497 512 BY SIMILARITY.  
 FT DISULFID 514 523 BY SIMILARITY.  
 FT DISULFID 530 541 BY SIMILARITY.  
 FT DISULFID 535 550 BY SIMILARITY.  
 FT DISULFID 552 561 BY SIMILARITY.  
 FT DISULFID 636 647 BY SIMILARITY.  
 FT DISULFID 641 656 BY SIMILARITY.  
 FT DISULFID 658 667 BY SIMILARITY.  
 FT DISULFID 674 685 BY SIMILARITY.  
 FT DISULFID 679 694 BY SIMILARITY.  
 FT DISULFID 696 705 BY SIMILARITY.  
 FT DISULFID 712 723 BY SIMILARITY.  
 FT DISULFID 717 732 BY SIMILARITY.  
 FT DISULFID 734 743 BY SIMILARITY.  
 FT DISULFID 751 762 BY SIMILARITY.  
 FT DISULFID 756 771 BY SIMILARITY.  
 FT DISULFID 773 782 BY SIMILARITY.  
 FT DISULFID 789 800 BY SIMILARITY.  
 FT DISULFID 794 809 BY SIMILARITY.  
 FT DISULFID 811 820 BY SIMILARITY.  
 FT DISULFID 827 838 BY SIMILARITY.  
 FT DISULFID 832 847 BY SIMILARITY.  
 FT DISULFID 849 858 BY SIMILARITY.  
 FT DISULFID 858 877 BY SIMILARITY.  
 FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 385 385 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 560 560 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 748 748 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 960 960 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 991 991 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1046 1046 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1242 AA; 135343 MW; 59557A4B7F052860 CRC64;

Query Match  
 Best local similarity

56.7%  
 50.0%

Score 38; DB 1; Length 1242;  
 Pred. No. 43;

Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DEKNSPECITLGP 12

Db 725 DEGDSFKLCSP 736

# RESULT 5

UBC7\_ARATH STANDARD; PRT; 166 AA.  
 AC Q42541; Q8RXQ3;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Ubiquitin-conjugating enzyme E2 7 (EC 6.3.2.19) (ubiquitin-protein  
 ligase 7) (ubiquitin carrier protein 7).  
 GN UBC7 OR A5G59300 OR NMCL7.21.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND FUNCTION.  
 RC STRAIN=cv. Columbia; TISSUE=seedling;  
 RX MEDLINE=96218124; PubMed=8647807;  
 RA van Nocker S., Walker J.M., Vierstra R.D.;  
 RT "The Arabidopsis thaliana UBC7/13/14 genes encode a family of  
 RT multiubiquitin chain-forming E2 enzymes."  
 RL J. Biol. Chem. 271:12150-12158 (1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=9156233; PubMed=10048488;  
 RA Asamizu E., Sato S., Kaneko T., Nakamura Y., Kotani H., Miyajima N.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VIII.  
 RT Sequence features of the regions of 1,081,958 bp covered by seventeen  
 RT physically assigned P1 and TAC clones."  
 RL DNA Res. 5:379-391 (1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=22954850; PubMed=14593172;  
 RA Yamada K., Lim U., Dale J.M., Chen H., Shinn P., Palm C.J.,  
 RA Boutinick A.M., Wu H.C., Kim C.T., Nguyen M., Pham P.K., Cheuk R.F.,  
 RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,  
 RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,  
 RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,  
 RA Arakawa T., Bahl J., Banno F., Bowser L., Brooks S.Y., Carrinci P.,  
 RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,  
 RA Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Kames N.M.,  
 RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai U.,  
 RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,  
 RA Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,  
 RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;  
 RT "Empirical analysis of transcriptional activity in the Arabidopsis  
 RT genome."  
 RL Science 302:842-846 (2003).  
 RN [4]  
 RP FUNCTION: Catalyzes the covalent attachment of ubiquitin to other  
 RP proteins so as to signal them for selective protein degradation.  
 CC Involved in the formation of multiubiquitin chains.  
 CC CATALYTIC ACTIVITY: ATP + ubiquitin + protein lysine = AMP +  
 CC diphosphate + protein-N-ubiquityllysine.  
 CC -1- PATHWAY: Ubiquitin conjugation; second step.  
 CC -1- MISCELLANEOUS: A cysteine residue is required for ubiquitin-  
 CC thioester formation (by similarity).  
 CC -1- SIMILARITY: Belongs to the ubiquitin-conjugating enzyme family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----  
 CC DR EMBL, U33757; AAC9321.1; -  
 CC DR EMBL, AB016890; BAB09775.1; -  
 CC DR EMBL, AY080733; AAL6003.1; ALT\_INIT.  
 CC DR EMBL, BT005798; AAO64200.1; ALT\_INIT.  
 CC DR PIR, S71209; S71209.  
 CC DR HSSP, Q02159; 2UC2.  
 CC DR InterPro, IPR000608; UBC\_conjugat.  
 CC DR Pfam, PF00179; UQ\_con; 1\_  
 CC DR ProDom, PD000461; UBC\_conjugat; 1.  
 CC DR SMART, SM00212; UBCc; 1.  
 CC DR PROSITE, PS00183; UBIQUITIN\_CONJUGAT\_1; 1.  
 CC DR PROSITE, PS00127; UBIQUITIN\_CONJUGAT\_2; 1.  
 CC KW Ubl conjugation pathway; Ligase; Multiubiquitin family.  
 CC FT BINDING 89 89 UBIQUITIN (BY SIMILARITY).  
 CC SQ SEQUENCE 166 AA; 18722 MW; 967F672ADB06891 CRC64;

Query Match 56.0%; Score 37.5; DB 1; Length 166;

Best Local Similarity 60.0%; Pred. No. 6.6;  
 Matches 9; Conservative 1; Mismatches 2; Indels 3; Gaps 1;

QY 1 DEKNSPE--CITGP 12  
 Db 30 DEKNFEMSVITLGP 44

# RESULT 6

UBC7\_ARATH STANDARD; PRT; 166 AA.  
 AC Q42541;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Ubiquitin-conjugating enzyme E2 13 (EC 6.3.2.19) (ubiquitin-protein  
 ligase 13) (ubiquitin carrier protein 13).  
 GN UBC13 OR A3G646460 OR P1815.180.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND FUNCTION.  
 RC STRAIN=cv. Columbia; TISSUE=seedling;  
 RX MEDLINE=96218124; PubMed=8647807;  
 RA van Nocker S., Walker J.M., Vierstra R.D.;  
 RT "The Arabidopsis thaliana UBC7/13/14 genes encode a family of  
 RT multiubiquitin chain-forming E2 enzymes."  
 RL J. Biol. Chem. 271:12150-12158 (1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=21016720; PubMed=11130713;  
 RA Salenoubat M., Lemcke K., Rieger M., Perez-Alonso M., Obermayer B.,  
 RA Fartmann B., Valle G., Bloecker H., Perez-Perez A., Othenwaelder B.,  
 RA Delany M., Boutry M., Grivell L.A., Maché R., Puigdomenech P.,  
 RA De Simone V., Chistone N., Artiguenave F., Robert C., Brottier P.,  
 RA Wincker P., Cattoilico L., Weissenbach J., Saurin W., Quetier F.,  
 RA Schaefer M., Mueller-Auer S., Gabl C., Fuchs M., Benes V.,  
 RA Wurmbach E., Dyzonek H., Erile H., Jordan N., Barraud S.,  
 RA Wiedelmann R., Krantz H., Voss H., Holland R., Barraud P., Nyakatura G.,  
 RA Vezzi A., D'Angelo M., Pallavicini A., Toppi S., Simionati B.,  
 RA Contrat A., Hornischer K., Kauer G., Boehmer T.-H., Nordstiek G.,  
 RA Reichel J., Scharte M., Schoen O., Barges M., Terol J., Climent J.,  
 RA Navarro P., Collado C., Perez-Perez A., Othenwaelder B., Duchemin D.,  
 RA Cooke R., Lande M., Berger-Liauro C., Purnelle B., Masuy D.,  
 RA de Haan M., Maeres A.C., Alcaraz J.-P., Cortet R., Casacubeta E.,  
 RA Montfort A., Argitrou A., Flores M., Ligouri R., Vitale D.,  
 RA Manhaupt G., Haase D., School H., Rudd S., Zaccaria P., Wewes H.-W.,  
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,



RA Rooney T., Rizzo M., Waite A., Utterback T., Fujii C.Y., Shea T.P.,  
 RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,  
 RA Pai G., Miltschker J., Sellers P., Gill J.E., Feldblyum T.V.,  
 RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,  
 RA Fraser C.M., Kameko T., Nakamura Y., Sato S., Kato T., Asanizu E.,  
 RA Sasamoto S., Kimura T., Ideawake K., Kawashima K., Kishida Y.,  
 RA Kiyokawa S., Kohara M., Matsumoto M., Matsumoto A., Muraki A.,  
 RA Nakayama S., Nakazaki N., Shino S., Takeuchi C., Wada T.,  
 RA Watanabe A., Yamada M., Yasuda M., Tabata S.,  
 RA "Sequence and analysis of chromosome 3 of the plant Arabidopsis  
 thaliana";  
 RT Nature 408:820-822(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=22954850; PubMed=14593172;  
 RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,  
 RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,  
 RA Karlin-Newmann G., Liu S.K., Lam B., Sakano H., Wu T., Yu G.,  
 RA Miranda M., Quach H.L., Tripp J.M., Chang C.H., Lee J.M., Toriani M.J.,  
 RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,  
 RA Arkwa T., Banh J., Banno F., Bowser L., Brooks S.V., Carninci P.,  
 RA Chao Q., Choy N., Ejima A., Goldsmith A.D., Gurjal M., Hansen N.F.,  
 RA Hayashizaki Y., Johnson-Hopson C., Huan V.W., Iida K., Karnes M.,  
 RA Khan S., Koseema E., Ishida J., Jiang P.X., Jones T., Kawai J.,  
 RA Kantuya A., Meyers R., Nakajima M., Narusaka M., Seki M., Sakurai T.,  
 RA Sarcu M., Tansie R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,  
 RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.,  
 RA "Empirical analysis of transcriptional activity in the Arabidopsis  
 genome";  
 RT Science 302:842-846(2003).  
 CC -1- FUNCTION: Catalyzes the covalent attachment of ubiquitin to other  
 proteins so as to signal them for selective protein degradation.  
 CC Involved in the formation of multiubiquitin chains.  
 CC -1- CATALYTIC ACTIVITY: ATP + ubiquitin + protein lysine = AMP +  
 diphosphate + protein N-ubiquityllysine.  
 CC -1- PATHWAY: Ubiquitin conjugation; second step.  
 CC -1- MISCELLANEOUS: A cysteine residue is required for ubiquitin-  
 thiolester formation (By similarity).  
 CC -1- SIMILARITY: Belongs to the ubiquitin-conjugating enzyme family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U33758; AAC49322.1; -;  
 DR EMBL; AL133298; CAB62037.1; -;  
 DR EMBL; AY050368; AAK91385.1; -;  
 DR EMBL; AY094040; AAM6196.1; -;  
 DR HSSP; Q02159; 2UC2.  
 DR InterPro; IPR000608; Ubc\_conjugat.  
 DR Pfam; PF00179; Ubc\_con; 1.  
 DR ProDom; PD000461; Ubc\_conjugat; 1.  
 DR SMART; SM00212; UBCG; 1.  
 DR PROSITE; PS00183; UBIQUITIN\_CONJUGAT\_1; 1.  
 DR PROSITE; PS00127; UBIQUITIN\_CONJUGAT\_2; 1.  
 KW Ubl conjugation pathway; ligase; Multigene family  
 FT BINDING 89 UBIQUITIN (By similarity).  
 FT BINDING 89  
 SQ SEQUENCE 166 AA; 18822 MW; DCC264242758275B CRC64;

Query Match 56.0%; Score 37.5; DB 1; Length 166;  
 Best Local Similarity 60.0%; Pred. No. 6.6;  
 Matches 9; Conservative 1; Mismatches 2; Indels 3; Gaps 1;

Qy 1 DEKNSFB---CITGP 12  
 |||||  
 Db 30 DEKNFMSVITIGP 44

RESULT 7  
 RB3 RAT  
 ID RB3 RAT STANDARD; PRT; 164 AA.  
 AC P35286;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ras-related protein Rab-13 (Fragment).  
 GN RAB13.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;  
 RX MEDLINE=92210533; PubMed=1313420;  
 RA Elferink L.A., Anzai K., Scheller R.H.,  
 RT "Rab15, a novel low molecular weight GTP-binding protein specifically  
 expressed in rat brain";  
 RL J. Biol. Chem. 267:5768-5775(1992).  
 CC -1- FUNCTION: May be involved in vesicular trafficking and  
 neurotransmitter release.  
 CC -1- TISSUE SPECIFICITY: Highest levels found in lung, kidney, whole  
 brain and spinal cord.  
 CC -1- SIMILARITY: Belongs to the small GTPase superfamily. Rab family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M83678; AAA41993.1; -;  
 DR PIR; D42148; D42148.  
 DR HSSP; P05713; 3RAB.  
 DR InterPro; IPR003579; GTPase\_Rab.  
 DR InterPro; IPR001806; Ras\_tmsf1rmg.  
 DR Pfam; PF00071; Ras; 1.  
 DR PRINTS; PR00449; RASTRNSFRNG.  
 DR SMART; SM00175; RAB; 1.  
 KW GTP-binding; Lipoprotein; Prenylation; Protein transport.  
 FT NON\_TER 1 1  
 FT NF\_BIND 24 28 GTP (By similarity).  
 FT NF\_BIND 62 85 GTP (By similarity).  
 FT DOMAIN <1 6 EFFECTOR REGION (By similarity).  
 FT LIPID 161 161 S-geranylgeranyl cysteine  
 (By similarity).  
 SQ SEQUENCE 164 AA; 18641 MW; A2DP9DC3ECCBA7A CRC64;

Query Match 55.2%; Score 37; DB 1; Length 164;  
 Best Local Similarity 63.6%; Pred. No. 8.1;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DEKNSFBCTIG 11  
 |||||  
 Db 154 DEKNSKCSIG 164

RESULT 8  
 PDXA\_HELPJ  
 ID PDXA\_HELPJ STANDARD; PRT; 307 AA.  
 AC Q9ZJ26;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 4-hydroxythreosine-4-phosphate dehydrogenase (EC 1.1.1.262) (4-  
 (phosphohydroxy)-L-threonine dehydrogenase).  
 GN PDXA OR JHP1490.  
 OS Helicobacter pylori J99 (Campylobacter pylori J99).  
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;



```

OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Molt D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guld B.C., deJonge B.L., Camel G.,
RA Tamano P.J., Caruso A., Uria-Nikelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
CC -1- FUNCTION: Catalyzes the NAD-dependent oxidation of 4-
CC (phosphohydroxy)-L-threonine (HTP) into 2-amino-3-oxo-4-
CC form 1-amino-3-(phosphohydroxy)propan-2-one (3-amino-2-oxopropyl
CC phosphate) (By similarity).
CC -1- CATALYTIC ACTIVITY: 4-(phosphonoxy)-L-threonine + NAD(+) = 2-
CC amino-3-oxo-4-phosphonoxybutyrate + NADH.
CC -1- PATHWAY: De novo synthesis of pyridoxine (vitamin B6) and
CC pyridoxal phosphate.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: Belongs to the pdxA family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: A5001570; AAD07064.1; -.
DR PIR: D71801; D71801.
DR HAMAP: MF_00536; -; 1.
DR InterPro: IPR005255; PdxA.
DR Pfam: PF04166; PdxA; 1.
DR TIGRPFAMs: TIGR00557; pdxA; 1.
DR Pyridoxine biosynthesis; Oxidoreductase; NAD; Complete proteome.
KW SEQUENCE 307 AA; 33706 MW; 77C5FD2DD9A71A61 CRC64;
SQ
Query Match 55.2%; Score 37; DB 1; Length 307;
Best Local Similarity 85.7%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 6 FECLIGP 12
DB 222 FECLIGP 228

```

```

RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.B., Gocayne J.D., Uterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori.";
RL Nature 388:539-547(1997).
CC -1- FUNCTION: Catalyzes the NAD-dependent oxidation of 4-
CC (phosphohydroxy)-L-threonine (HTP) into 2-amino-3-oxo-4-
CC (phosphohydroxy)butyric acid which spontaneously decarboxylate to
CC form 1-amino-3-(phosphohydroxy)propan-2-one (3-amino-2-oxopropyl
CC phosphate) (By similarity).
CC -1- CATALYTIC ACTIVITY: 4-(phosphonoxy)-L-threonine + NAD(+) = 2-
CC amino-3-oxo-4-phosphonoxybutyrate + NADH.
CC -1- PATHWAY: De novo synthesis of pyridoxine (vitamin B6) and
CC pyridoxal phosphate.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: Belongs to the pdxA family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB000655; AAD08621.1; -.
DR PIR: G64717; G64717.
DR TIGR: HP1583; -.
DR HAMAP: MF_00536; -; 1.
DR InterPro: IPR005255; PdxA.
DR Pfam: PF04166; PdxA; 1.
DR TIGRPFAMs: TIGR00557; pdxA; 1.
DR Pyridoxine biosynthesis; Oxidoreductase; NAD; Complete proteome.
KW SEQUENCE 307 AA; 33578 MW; B122B0A0B800D70E CRC64;
SQ
Query Match 55.2%; Score 37; DB 1; Length 307;
Best Local Similarity 85.7%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 6 FECLIGP 12
DB 222 FECLIGP 228

```

```

RESULT 9
PDXA_HELPY
ID_PDXA_HELPY STANDARD; PRT; 307 AA.
AC 026103;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 4-hydroxythreonine-4-phosphate dehydrogenase (EC 1.1.1.262) (4-
DE (phosphohydroxy)-L-threonine dehydrogenase).
GN PDXA OR HP1583.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Krelavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,

```

```

RESULT 10
TRBM_HUMAN
ID_TRBM_HUMAN STANDARD; PRT; 575 AA.
AC P07204;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Thrombomodulin precursor (fetomodulin) (TM) (CD141 antigen).
GN THBD OR TRBM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88004395; PubMed=2820710;
RX Suzuki K., Kusumoto H., Deyashiki Y., Nishioka J., Maruyama I.,
RX Zuehl M., Kawahara S., Honda G., Yamamoto S., Horiguchi S.;
RT "Structure and expression of human thrombomodulin, a thrombin
RT receptor on endothelium acting as a cofactor for protein C
RT activation.";
RL EMBL J. 6:11891-1897(1987).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88024950; PubMed=2822087;
RX Wen D., Dittman W.A., Ye R.D., Deaven L.L., Majerus P.W., Sadler J.E.;

```

- RT "Human thrombomodulin: complete cDNA sequence and chromosome  
localization of the gene." ;  
RL Biochemistry 26:4350-4357(1987).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87317665; PubMed=2819876;  
RA Jackson R.W., Beeler D.L., Fritze L., Soff G., Rosenberg R.D.;  
RT "Human thrombomodulin gene is intron depleted: nucleic acid sequences  
of the cDNA and gene predict protein structure and suggest sites of  
regulatory control." ;  
RL Proc. Natl. Acad. Sci. U.S.A. 84:6425-6429(1987).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88227901; PubMed=2836377;  
RA Shirai T., Shiojiri S., Ito H., Yamamoto S., Kusumoto H.,  
RA Dayashiki Y., Maruyama I., Suzuki K.;  
RT "Gene structure of human thrombomodulin, a cofactor for thrombin-  
catalyzed activation of protein C." ;  
RL J. Biochem. 103:281-285(1988).  
RN [5]  
RP SEQUENCE FROM N.A.  
RA Rieder M.J., Amel T.Z., Carrington D.P., Chung M.-W., Lee K.L.,  
RA Ouna M., Poel C.L., Toth E.J., Yi O., Nickerson D.A.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21638749; PubMed=11780052;  
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,  
RA Bailey J., Barlow K.F., Bates K.N., Beard L.W., Beare D.M.,  
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,  
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,  
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,  
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
RA Levenshio M.H., Levertha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
RA Marsh V.L., Martin S.L., McDonach L.J., McLeay K., McMurtry A.A.,  
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
RA Rice C.M., Ross M.T., Scott C.E., Senra H.K., Showkeen R., Sims S.,  
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
RA Swann R.M., Symmore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
RA Whitehead S.L., Whitteker P., Willey D.L., Williams L., Williams S.A.,  
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
Rogers J.;  
RT "The DNA sequence and comparative analysis of human chromosome 20." ;  
RL Nature 414:865-871(2001).  
RN [7]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shanen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,  
RA Datchenko L., Waisina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ueda T.B., Tohshiyuki S., Carinici P., Prange C.,  
RA Rata S.S., Loguailano N.A., Peters G.J., Abramson R.D., Mulhally S.J.,  
RA Bosak S.A., McMan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.O., Hulyk S.W.,  
RA Villalón D.K., Wuzry D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahney J., Helton E., Keltman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences." ;  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [8]  
RP CARBOHYDRATE-LINKAGE SITE SER-492, AND MUTAGENESIS.  
RX MEDLINE=94029900; PubMed=8216207;  
RA Grunewald B., Hassell T., Vlahos C.J., Parkinson J.F., Bang N.U.,  
RA Grunewald B.W.;  
RT "Identification of the predominant glycosaminoglycan-attachment site  
in soluble recombinant human thrombomodulin: potential regulation of  
functionality by glycosyltransferase competition for serine474." ;  
RL Biochem. J. 295:131-140(1993).  
RN [9]  
RP STRUCTURE BY NMR OF 389-407.  
RX MEDLINE=96007474; PubMed=7559494;  
RA Adler M., Seto M.H., Nitecki D.E., Lin J.H., Light D.R., Morser J.;  
RT "The structure of a 19-residue fragment from the C-loop of the fourth  
epidermal growth factor-like domain of thrombomodulin." ;  
RL J. Biol. Chem. 270:23366-23372(1995).  
RN [10]  
RP STRUCTURE BY NMR OF 364-407.  
RX MEDLINE=96100636; PubMed=8528067;  
RA Weininger D.P., Hunter M.U., Komives E.A.;  
RT "Synthesis, activity, and preliminary structure of the fourth  
EGF-like domain of thrombomodulin." ;  
RL Protein Sci. 4:1683-1695(1995).  
RN [11]  
RP STRUCTURE BY NMR OF 427-444.  
RX MEDLINE=95034791; PubMed=7947766;  
RA Srivivasan J., Hu S., Hrabal R., Zhu Y., Komives E.A., Ni F.;  
RT "Thrombin-bound structure of an EGF subdomain from human  
thrombomodulin determined by transferred nuclear Overhauser  
effects." ;  
RL Biochemistry 33:13553-13560(1994).  
RN [12]  
RP STRUCTURE BY NMR OF 427-444.  
RX MEDLINE=96276211; PubMed=8745396;  
RA Hrabal R., Komives E.A., Ni F.;  
RT "Structural resiliency of an EGF-like subdomain bound to its target  
protein, thrombin." ;  
RL Protein Sci. 5:195-203(1996).  
RN [13]  
RP STRUCTURE BY NMR OF 405-444.  
RX MEDLINE=98035729; PubMed=9367781;  
RA Sampoll Bentler B.A., Hunter M.J., Weininger D.P., Komives E.A.;  
RT "Structure of the fifth EGF-like domain of thrombomodulin: an  
EGF-like domain with a novel disulfide-bonding pattern." ;  
RL J. Mol. Biol. 273:913-926(1997).  
RN [14]  
RP VARIANT TED TYR-486.  
RX MEDLINE=9511115; PubMed=7811889;  
RA Oehlén A.-K., Marlar R.A.;  
RT "The first mutation identified in the thrombomodulin gene in a  
45-year-old man presenting with thromboembolic disease." ;  
RL Blood 85:330-336(1995).  
RN [15]  
RP VARIANT TED TYR-486, AND VARIANTS THR-43; ALA-79; SER-495 AND LEU-501.  
RX MEDLINE=97241986; PubMed=9198186;  
RA Oehlén A.-K., Norlund L., Marlar R.A.;  
RT "Thrombomodulin gene variations and thromboembolic disease." ;  
RL Thromb. Haemost. 78:396-400(1997).  
RN [16]  
RP VARIANT VAL-473.  
RX MEDLINE=97206518; PubMed=9157575;  
RA Norlund L., Holm J., Zoller B., Oehlén A.-K.;  
RT "A common thrombomodulin amino acid dimorphism is associated with  
myocardial infarction." ;  
RL Thromb. Haemost. 77:248-251(1997).  
RN [17]  
RP VARIANT THR-43.  
RX MEDLINE=99057299; PubMed=9843165;

```

RA Doegen C.J.M., Kunz G., Rosendal F.R., Lane D.A., Vos H.L.,
RA Stubbs P.J., Manger Cats V., Ireland H., 127G to A coding for Ala25Thr,
RT "A mutation in the thrombomodulin gene, 127G to A coding for Ala25Thr,
RT and the risk of myocardial infarction in men.";
RL Thromb. Haemost. 80:743-748(1998).
RN [18]
RP VARIANT VAL-473.
RX MEDLINE=21143723; PubMed=11245641.
RA Wu K.K., Alekic N., Alm C., Boerwinkle E., Folsom A.R.,
RA Juneja H.;
RT "Thrombomodulin Ala45Val polymorphism and risk of coronary heart
RT disease.";
RL Circulation 103:1386-1389(2001).
RN [19]
RP VARIANT TED TYR-486, AND VARIANT VAL-473.
RX MEDLINE=22135346; PubMed=12139752;
RA Faloni E.M., Franchi F., Castaman G., Biguzzi E., Rodeghiero F.;
RT "Mutations in the thrombomodulin gene are rare in patients with
RT severe thrombophilia.";
RL Br. J. Haematol. 118:595-599(2002).
CC -1- FUNCTION: Thrombomodulin is a specific endothelial cell receptor
CC that forms a 1:1 stoichiometric complex with thrombin. This
CC complex is responsible for the conversion of protein C to the
CC activated protein C (protein Ca). Once evolved, protein Ca
CC scissions the activated cofactors of the coagulation mechanism,
CC factor Va and factor VIIIa, and thereby reduces the amount of
CC thrombin generated.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: Endothelial cells are unique in synthesizing
CC thrombomodulin.

Query Match 55.2%; Score 37; DB 1; Length 575;
Best Local Similarity 75.0%; Pred. No. 30;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 5 SPECIFIC 12
Db 461 TPECICGP 468

RESULT 11
TRBM_MOUSE
ID TRBM_MOUSE STANDARD; PRT; 577 AA.
AC P15306;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thrombomodulin precursor (Fetomodulin) (TM).
GN THBD.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
CX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89128454; PubMed=2536925;
RA Dittman W.A., Majerus P.W.;
RT "Sequence of a cDNA for mouse thrombomodulin and comparison of the
RT predicted mouse and human amino acid sequences.";
RL Nucleic Acids Res. 17:802-802(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89008498; PubMed=2844823;
RA Dittman W.A., Kumada T., Sadler J.E., Majerus P.W.;
RT "The structure and function of mouse thrombomodulin. Phorbol
RT myristate acetate stimulates degradation and synthesis of
RT thrombomodulin without affecting mRNA levels in hemangioendothelial
RL J. Biol. Chem. 263:15815-15822(1988).
CC -1- FUNCTION: Thrombomodulin is a specific endothelial cell receptor
CC that forms a 1:1 stoichiometric complex with thrombin. This
CC complex is responsible for the conversion of protein C to the
CC activated protein C (protein Ca). Once evolved, protein Ca
CC scissions the activated cofactors of the coagulation mechanism,
CC scissions the activated cofactors of the coagulation mechanism,

```

CC	factor Va and factor VIIIa, and thereby reduces the amount of
CC	thrombin generated.
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.
CC	-1- TISSUE SPECIFICITY: Endothelial cells are unique in synthesizing
CC	thrombomodulin (By similarity).
CC	-1- SIMILARITY: Contains 6 EGF-like domains.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; X14432; CAA32597.1; .
DR	PIR; S08488; A60501.
DR	HSSP; P07204; IEGT.
DR	MGD; MGI:98736; Thbd.
DR	GO; GO:0009790; P:embryonic development; IMP.
DR	GO; GO:0007565; P:pregnancy; IMP.
DR	InterPro; IPR000152; Asx_hydroxyl_5.
DR	InterPro; IPR001881; EGF_Ca.
DR	InterPro; IPR006209; EGF-like.
DR	InterPro; IPR001304; Lectin C.
DR	InterPro; IPR001491; Thrombomodulin.
DR	Pfam; PF00008; EGF_3.
DR	Pfam; PF00059; Lectin_c; 1.
DR	PRINTS; PR00907; THROMBOMODULIN.
DR	SMART; SM00034; CLECT; 1.
DR	SMART; SM00179; EGF_Ca; 1.
DR	PROSITE; PS00010; ASX_HYDROXYL_2.
DR	PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
DR	PROSITE; PS00022; EGF_1; FALSE_NEG.
DR	PROSITE; PS01166; EGF_2; 3.
DR	PROSITE; PS50026; EGF_3; 3.
DR	PROSITE; PS01167; EGF_CA; 2.
KM	Endothelial cell; Receptor; blood coagulation; Repeat; Transmembrane;
KM	Glycoprotein; Signal; EGF-like domain.
FT	SIGNAL 1 16 POTENTIAL.
FT	CHAIN 17 577 THROMBOMODULIN.
FT	DOMAIN 17 517 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 518 541 POTENTIAL.
FT	DOMAIN 542 577 CYTOPLASMIC (POTENTIAL).
FT	DOMAIN 240 280 EGF-LIKE 1.
FT	DOMAIN 283 323 EGF-LIKE 2.
FT	DOMAIN 324 362 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN 364 404 EGF-LIKE 4.
FT	DOMAIN 403 439 EGF-LIKE 5.
FT	DOMAIN 440 480 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT	DISULFID 244 255 BY SIMILARITY.
FT	DISULFID 251 264 BY SIMILARITY.
FT	DISULFID 266 279 BY SIMILARITY.
FT	DISULFID 287 295 BY SIMILARITY.
FT	DISULFID 291 307 BY SIMILARITY.
FT	DISULFID 309 322 BY SIMILARITY.
FT	DISULFID 328 339 BY SIMILARITY.
FT	DISULFID 335 348 BY SIMILARITY.
FT	DISULFID 350 361 BY SIMILARITY.
FT	DISULFID 368 377 BY SIMILARITY.
FT	DISULFID 373 387 BY SIMILARITY.
FT	DISULFID 389 403 BY SIMILARITY.
FT	DISULFID 407 416 BY SIMILARITY.
FT	DISULFID 412 424 BY SIMILARITY.
FT	DISULFID 426 438 BY SIMILARITY.
FT	DISULFID 444 454 BY SIMILARITY.
FT	DISULFID 449 463 BY SIMILARITY.
FT	DISULFID 465 479 BY SIMILARITY.
FT	CARBOHYD 113 113 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 243 243 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 256 256 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 494 494 O-LINKED (GLYCOSAMINOGLYCAN) (BY

```

FT      SEQUENCE      577 AA; 61867 MW; B20E50B0FE745014 CRC64;
SQ      QUERY MATCH      55.2%; Score 37; DB 1; Length 577;
      Best Local Similarity 75.0%; Pred. No. 30;
      Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY      5 SPECILGP 12
      Db 460 SYECICGP 467

RESULT 12
PD13 MOUSE
ID PD13 MOUSE STANDARD; PRT; 664 AA.
AC Q92184;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Protein-arginine deiminase type III (EC 3.5.3.15) (peptidylarginine
DE deiminase III).
GN PAD13 OR PD13.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Epidermis; PubMed=10092850;
RX MEDLINE=9912810; PubMed=178991;
RA Rud A.A., Ikejiri Y., Ono H., Yonekawa T., Shiraiwa M., Kawada A.,
RA Takahara H.;
RT "Molecular cloning of cDNAs of mouse peptidylarginine deiminase type
RT I, type III and type IV, and the expression pattern of type I in
RT mouse."
RT Eur. J. Biochem. 259:660-669 (1999).
RN [2]
RP CHARACTERIZATION, AND TISSUE SPECIFICITY.
RX MEDLINE=9213676; PubMed=178991;
RA Terakawa H., Takahara H., Sugawara K.;
RT "Three types of mouse peptidylarginine deiminase: characterization and
RT tissue distribution."
RT J. Biochem. 110:661-666 (1991).
CC -1- FUNCTION: Catalyzes the deamination of arginine residues of
CC proteins.
CC -1- CATALYTIC ACTIVITY: Protein L-arginine + H(2)O = protein L-
CC citrulline + NH(3).
CC -1- COFACTOR: Requires calcium ions.
CC -1- TISSUE SPECIFICITY: Epidermis and hair follicles.
CC -1- SIMILARITY: Belongs to the protein arginine deiminase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB013849; BAA34182.1; -
DR MGD; MGI:133891; Pad13.
DR InterPro: IPR004303; Prot_arg_deim.
DR Pfam; PF03068; PAD, 1.
KW Hydroxylase; Calcium-binding; Multigene family.
FT CA BIND 504 515 EF-HAND (POTENTIAL).
SQ SEQUENCE 664 AA; 75099 MW; BB005AE930AE6F31 CRC64;

Query Match 55.2%; Score 37; DB 1; Length 664;
Best Local Similarity 41.7%; Pred. No. 34;
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY      1 DEKNSFECLGP 12
      Db 460 SYECICGP 467

RESULT 13
DL14 MOUSE
ID DL14 MOUSE STANDARD; PRT; 686 AA.
AC Q9J7L1; Q9J7L2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Delta-like protein 4 precursor (Drosophila Delta homolog 4).
GN Dlt4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Adipose tissue;
RX MEDLINE=20296719; PubMed=10837024;
RA Shutter J.R., Scully S., Fan W., Richards W.G., Kitajewski J.,
RA Delandre G.A., Kintner C.R., Stark K.L.;
RT "Dlt4, a novel Notch ligand expressed in arterial endothelium."
RT Genes Dev. 14:1313-1318 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21064937; PubMed=1134954;
RA Yoneya T., Tanaka T., Nagao K., Yamada Y., Yamamoto T., Miyatani S.,
RA Nishikawa M.;
RT "Molecular cloning of Delta-4, a new mouse and human Notch ligand."
RT J. Biochem. 129:27-34 (2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=129/Ola.
RA Malinos C., Lewis J., Ish-Horowitz D.;
RT "A novel Notch ligand expressed in embryonic and tumour
RT vasculature."
RT Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: Plays a role in the Notch signaling pathway. Activates
CC Notch-1 and Notch-4.
CC -1- SUBUNIT: Binds to Notch-1 and Notch-4.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC -1- TISSUE SPECIFICITY: Expressed in vascular endothelium.
CC -1- SIMILARITY: Contains 8 EGF-like domains.
CC -1- SIMILARITY: Contains 1 DSL domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF253469; AAF76428.1; -
DR EMBL; AF273454; AAF78785.1; -
DR EMBL; AB043893; BAB18580.1; -
DR HSSP; P08709; IBF9.
DR MGD; MGI:1859388; Dlt4.
DR GO; GO:0005112; F:Notch binding, IDA.
DR InterPro: IPR000152; Asx_hydroxyl_S.
DR InterPro: IPR001774; DSL.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001861; EGF_Ca.
DR InterPro: IPR001438; EGF_II.
DR InterPro: IPR006209; EGF_III.
DR InterPro: IPR002049; Laminin_EGF.
DR Pfam; PF01414; DSL, 1.
DR Pfam; PF00008; EGF, 7.
DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR00011; EGFBL00N.
DR SMART; SM00051; DSL, 1.
DR SMART; SM00179; EGF_CA, 3.

```

```

DR PROSITE; PS00010; ASX HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 8.
DR PROSITE; PS01186; EGF_2; 7.
DR PROSITE; PS00026; EGF_3; 8.
KW Signal; EGF-like domain; Repeat; Transmembrane; Developmental protein;
KM Differentiation; Glycoprotein.
FT SIGNAL 1 26
FT CHAIN 27 686
FT DOMAIN 27 532
FT TRANSMEM 533 553
FT DOMAIN 554 686
FT DOMAIN 156 218
FT DOMAIN 219 252
FT DOMAIN 253 323
FT DOMAIN 323 323
FT DOMAIN 323 361
FT DOMAIN 364 401
FT DOMAIN 403 439
FT DOMAIN 441 477
FT DOMAIN 481 519
FT DOMAIN 223 234
FT DISULFID 227 240
FT DISULFID 242 251
FT DISULFID 254 265
FT DISULFID 260 271
FT DISULFID 273 282
FT DISULFID 289 301
FT DISULFID 295 311
FT DISULFID 313 322
FT DISULFID 329 340
FT DISULFID 334 349
FT DISULFID 351 360
FT DISULFID 367 378
FT DISULFID 372 389
FT DISULFID 391 400
FT DISULFID 407 418
FT DISULFID 412 427
FT DISULFID 429 438
FT DISULFID 445 465
FT DISULFID 467 476
FT DISULFID 485 496
FT DISULFID 490 507
FT DISULFID 509 518
FT CARBOHYD 79 79
FT CARBOHYD 109 109
FT CARBOHYD 162 162
FT CARBOHYD 297 297
FT CARBOHYD 394 394
FT CONFLICT 419 419
FT CONFLICT 527 528
FT CONFLICT 627 627
FT CONFLICT 639 649
SQ SEQUENCE 686 AA; 75005 MW; 7C204FD3B087EB41 CRC64;

Query Match 55.2%; Score 37; DB 1; Length 686;
Best Local Similarity 41.7%; Pred. No. 36;
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 DEKNSFECLICP 12
DB 342 DQENSYHCLICP 353

```

```

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderiinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN (1)
RN SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Kirsten J.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RN REVISIONS.
RA Waterston R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; U10402; AAA19070.2; -.
CC PIR; T15765; T15765.
CC WormPep; C34E10.8; C34E10.8.
CC Hypothetical protein.
KM SEQUENCE 1024 AA; 112206 MW; AC75AFB9370A757E CRC64;

Query Match 55.2%; Score 37; DB 1; Length 1024;
Best Local Similarity 70.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 DEKNSFECLICP 10
DB 330 DEKNSITCYL 339

RESULT 15
ID CTRO_HUMAN STANDARD; PRT; 1286 AA.
AC 014578; Q0UP27;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytro protein (Rho-interacting, serine/threonine kinase 21)
DE (Fragment)
DE CIT OR STK21 OR KIAA0949.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RA Connell M., Goela D., Harper M.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE OF 347-1286 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99246063; PubMed=10231032;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 6:63-70(1999).
CC -!- FUNCTION: Putative RHO/RAC effector that binds to the GTP-bound
CC forms of RHO and RAC1. It probably binds p21 with a tighter
CC specificity in vivo (by similarity).
CC -!- SIMILARITY: Contains 1 CNH domain.
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 1 zinc-dependent phospho-ester and DAG
CC binding domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AC002563; AAB71327.1; -  
 DR EMBL; AB023166; BAA76793.1; -  
 DR Genew; HGNC.1985; CIT.  
 DR MTM; 605629; -  
 DR InterPro; IPR001180; Citron.  
 DR InterPro; IPR002219; DAG\_PE-bind.  
 DR InterPro; IPR001849; PH.  
 DR InterPro; IPR000861; REM\_repeat.  
 DR Pfam; PF00780; CNH; 1.  
 DR Pfam; PF00130; DAG\_PE-bind; 1.  
 DR Pfam; PF00169; PH; 1.  
 DR SMART; SM00109; CL; 1.  
 DR SMART; SM00036; CNH; 1.  
 DR SMART; SM00233; PH; 1.  
 DR PROSITE; PS00479; DAG\_PE\_BIND\_DOM\_1; 1.  
 DR PROSITE; PS00081; DAG\_PE\_BIND\_DOM\_2; 1.  
 DR PROSITE; PS50003; PH\_DOMAIN; 1.  
 KM Coiled coil; Phorbol-ester binding; SH3-binding.  
 FT NON TER 1 1  
 FT DOMAIN <1 558 COILED COIL (POTENTIAL).  
 FT DOMAIN 365 561 RHO/RAC BINDING.  
 FT DOMAIN 509 512 POLY-LYS.  
 FT DOMAIN 622 670 PHORBOL-ESTER AND DAG BINDING.  
 FT DOMAIN 702 822 PH.  
 FT DOMAIN 851 1148 CNH.  
 FT SITE 1212 1217 SH3-BINDING (POTENTIAL).  
 SQ SEQUENCE 1286 AA; 146506 MW; 498101F79EA75E85 CRC64;

Query Match 55.2%; Score 37; DB 1; Length 1286;  
 Best Local Similarity 75.0%; Pred. No. 68;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 DEKNSPEC 8  
 |||: |||  
 Db 326 DEKNSPEC 333

Search completed: March 22, 2004, 06:52:58  
 Job time: 1.46026 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:39:53 ; Search time 2.4167 Seconds  
(without alignments)  
1568.003 Million cell updates/sec

Title: US-09-662-293-5  
Perfect score: 67  
Sequence: 1 DEKNSFECLGP 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL.25:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp rodent:\*  
12: sp.virus:\*  
13: sp.vertibrate:\*  
14: sp.unclassified:\*  
15: sp.virus:\*  
16: sp.bacteriap:\*  
17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	67	100.0	555	Q9U6R7	Q9U6R7 dermatophag
2	47	70.1	180	Q9TXK2	Q9TXK2 caenorhabdi
3	44	65.7	333	Q81G84	Q81G84 bacillus ce
4	42	62.7	181	Q17727	Q17727 caenorhabdi
5	42	62.7	258	Q89Q37	Q89Q37 bradyrhizob
6	41	61.2	291	Q9PR76	Q9PR76 ureaplasma
7	40	59.7	39	Q56605	Q56605 vibrio chol
8	40	59.7	67	Q8XY55	Q8XY55 raietonia s
9	40	59.7	86	Q8YU00	Q8YU00 beak and fe
10	40	59.7	236	Q8NMNO	Q8NMNO corynebacte
11	40	59.7	289	Q8DV23	Q8DV23 streptococc
12	40	59.7	372	Q87U11	Q87U11 vibrio para
13	40	59.7	403	Q9KLY9	Q9KLY9 vibrio chol
14	39.5	59.0	455	Q7UH40	Q7UH40 rhodospirill
15	39	58.2	187	Q26593	Q26593 methanobact
16	39	58.2	233	Q86BK7	Q86BK7 drosophila

17	39	58.2	234	16	Q8YRX1	Q8YRX1 anabaena sp
18	39	58.2	248	17	Q9YCF6	Q9YCF6 aeropyrum p
19	39	58.2	284	16	Q7VZU1	Q7VZU1 prochloroco
20	39	58.2	285	16	Q897F5	Q897F5 clostridium
21	39	58.2	325	2	Q05698	Q05698 rhizobium 1
22	39	58.2	326	2	P95623	P95623 rhizobium 1
23	39	58.2	384	2	Q30659	Q30659 rhizobium 1
24	39	58.2	528	5	Q9U1R8	Q9U1R8 caenorhabdi
25	39	58.2	2531	5	Q16004	Q16004 lytechinus
26	39	58.2	4998	11	Q8CG65	Q8CG65 mus musculu
27	39	58.2	5251	5	Q81TD4	Q81TD4 plasmodium
28	38.5	57.5	497	13	Q80311	Q80311 brachydanio
29	38	56.7	119	13	Q8BFL6	Q8BFL6 ginglymesto
30	38	56.7	141	16	Q9Z8U1	Q9Z8U1 chlamydia p
31	38	56.7	155	11	Q994S2	Q994S2 mus musculu
32	38	56.7	235	11	Q8K3A4	Q8K3A4 mus musculu
33	38	56.7	242	7	P79609	P79609 ambystoma m
34	38	56.7	264	2	Q54390	Q54390 microcystis
35	38	56.7	264	2	Q54391	Q54391 microcystis
36	38	56.7	283	16	Q8BR33	Q8BR33 pseudomonas
37	38	56.7	305	12	Q9YTL5	Q9YTL5 melanoplus
38	38	56.7	334	16	Q9S2C6	Q9S2C6 streptomyce
39	38	56.7	346	7	P79458	P79458 ambystoma m
40	38	56.7	369	10	Q9F210	Q9F210 phaseolus v
41	38	56.7	380	10	Q04978	Q04978 oryza sativ
42	38	56.7	397	2	Q8KPD2	Q8KPD2 gamma-prote
43	38	56.7	501	8	Q8MDH2	Q8MDH2 ascaridia fa
44	38	56.7	587	5	Q9Y012	Q9Y012 plasmodium
45	38	56.7	664	10	Q7XX21	Q7XX21 oryza sativ

## ALIGNMENTS

RESULT 1  
Q9U6R7 PRELIMINARY; PRT; 555 AA.  
AC Q9U6R7;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE 98kDa HDM allergen  
OS Dermatophagoides farinae (House-dust mite).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
OC Acariformes; Sarcophitiformes; Astigmata; Psoroptida; Analgidae;  
OC Pyroglyphidae; Dermatophagoides.  
CX NCBI\_TaxID=6954;  
RV [1]  
RP SEQUENCE FROM N.A.  
RA Weber E.R., Hunter S., Stedman K., McCall C.;  
RT "Cloning and Characterization of a 98 kDa Allergen from  
RT Dermatophagoides farinae";  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF18772; AAD52672.1;  
DR GO; GO:0005576; C:extracellular; IEA.  
DR GO; GO:0008061; F:chitin binding; IEA.  
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR GO; GO:0006030; P:chitin metabolism; IEA.  
DR InterPro; IPR002557; Chitin\_bind\_perh.  
DR InterPro; IPR001223; Glyco\_hydro\_18.  
DR InterPro; IPR001579; Glyco\_hydro\_18A.  
DR Pfam; PF00704; Glyco\_hydro\_18; 1.  
DR ProDom; PD000471; Glyco\_hydro\_18; 1.  
DR SMART; SM00494; ChEBD2; 1.  
DR SMART; SM00636; Glyco\_18; 1.  
DR PROSITE; PS01095; CHITINASE\_18; 1.  
DR GlycoSite; Hydrolyase.  
SQ SEQUENCE 555 AA; 63238 MW; 0E4564A1A459B30B CRC64;  
KW  
Query Match 100.0%; Score 67; DB 5; Length 555;  
Best Local Similarity 100.0%; Pred. No. 0.00025;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DEKNSPECILGP 12  
 Db 404 DEKNSPECILGP 415

## RESULT 2

Q9TXK2 PRELIMINARY; PRT; 180 AA.

DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN F46F5.15.

OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;

RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;

RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).

## RESULT 3

Q81G84 PRELIMINARY; PRT; 333 AA.

DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE 2-aminocetylphosphonate transport ATP-binding protein pnt.  
 GN BC1324.

OS Bacillus cereus (strain ATCC 14579 / DSM 31).  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=226900;

RP SEQUENCE FROM N.A.  
 RC MEDLINE=22608415; PubMed=12721630;  
 RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,  
 RA Kapareli V., Bhattacharya A., Resnik G., Mikhailova N., Lapidas A.,  
 RA Chu L., Mazur M., Goltzman E., Larsen N., D'Souza M., Walunas T.,  
 RA Grechkin Y., Fusch G., Haseikorn R., Fonstein M., Ehrlich S.D.,

Qy 3 KNSPECILGP 12  
 Db 124 KNSPECILGP 133

## RESULT 3

Q81G84 PRELIMINARY; PRT; 333 AA.

DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE 2-aminocetylphosphonate transport ATP-binding protein pnt.  
 GN BC1324.

OS Bacillus cereus (strain ATCC 14579 / DSM 31).  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=226900;

RP SEQUENCE FROM N.A.  
 RC MEDLINE=22608415; PubMed=12721630;  
 RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,  
 RA Kapareli V., Bhattacharya A., Resnik G., Mikhailova N., Lapidas A.,  
 RA Chu L., Mazur M., Goltzman E., Larsen N., D'Souza M., Walunas T.,  
 RA Grechkin Y., Fusch G., Haseikorn R., Fonstein M., Ehrlich S.D.,

RA Overbeek R., Kyrpides N.,  
 RT "Genome sequence of Bacillus cereus and comparative analysis with  
 RT Bacillus anthracis.";  
 RL Nature 423:87-91(2003).

DR EMBL; AE017002; AAP08307.1; -  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.  
 DR GO; GO:0000166; F:nucleotide binding; IEA.  
 DR GO; GO:0006810; P:transport; IEA.

DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR SMART; SM00382; AAA; 1.

DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
 KW ATP-binding; Complete proteome.  
 SQ SEQUENCE 333 AA; 37238 MW; 762C5CB2A582BED CRC64;

Query Match 65.7%; Score 44; DB 16; Length 333;  
 Best Local Similarity 63.6%; Pred. No. 4;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EKNSPECILGP 12  
 Db 28 EKNSPECILGP 38

## RESULT 4

O17727 PRELIMINARY; PRT; 181 AA.

DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
 DE D1086.2 protein.

OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;

RP SEQUENCE FROM N.A.  
 RA Smye R.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology.";  
 RL Science 282:2012-2018(1998).

DR EMBL; Z81491; CAB04020.1; -  
 DR PIR; T20324; T20324.  
 DR WormPep; D1086.2; CE09060.  
 SQ SEQUENCE 181 AA; 20998 MW; 21698FF9D47BECA CRC64;

Query Match 62.7%; Score 42; DB 5; Length 181;  
 Best Local Similarity 72.7%; Pred. No. 5.4;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DEKNSPECILGP 11  
 Db 52 DEKNSPEMARG 62

## RESULT 5

Q89Q37 PRELIMINARY; PRT; 258 AA.

DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)



```

DE ABC transporter ATP-binding protein.
GN BLR3293.
OC Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22464998; PubMed=12597275;
RA Kaneo T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Saenamoto S., Watanabe A., Ideawa K., Iriuchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005947; BAC48558.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
DR ATP-binding; Complete proteome.
SQ SEQUENCE 258 AA; 28777 MW; 8FD39ADFB2475C7 CRC64;

Query Match 62.7%; Score 42; DB 16; Length 258;
Best Local Similarity 63.6%; Pred. No. 7.6;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EKNSFECTILGP 12
DB 33 EANDFVCLGP 43

RESULT 6
Q9PR76 PRELIMINARY; PRT; 291 AA.
ID Q9PR76
AC Q9PR76
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Iron(III) dicitrate transport ATP-binding protein.
GN PFCB OR U0069.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sevovar 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
RA Cassell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
RT urealyticum."
RL Nature 407:757-762(2000).
DR EMBL; AE002106; AAF30474.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SMO0382; AAA_1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW Complete proteome.
SQ SEQUENCE 291 AA; 33408 MW; 02FDB8FF9828E34C CRC64;

```

```

Query Match 61.2%; Score 41; DB 16; Length 291;
Best Local Similarity 72.7%; Pred. No. 13;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EKNSFECTILGP 12
DB 53 KXNSFTTLGP 63

RESULT 7
Q56605 PRELIMINARY; PRT; 39 AA.
ID Q56605
AC Q56605
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE IIVI (Fragment).
GN IIVI.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96414469; PubMed=8817490;
RA Camilli A., Mekalanos J.J.;
RT "Use of recombinase gene fusions to identify Vibrio cholerae genes
RT induced during infection."
RL Mol. Microbiol. 18:671-683(1995).
DR EMBL; U25729; AAC8357.1; -.
DR PIR; S70808; S70808.
FT NON_TER 1
FT NON_TER 39
SQ SEQUENCE 39 AA; 3967 MW; B0468F7DDBBA59 CRC64;

Query Match 59.7%; Score 40; DB 2; Length 39;
Best Local Similarity 63.6%; Pred. No. 3;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 EKNSFECTILGP 12
DB 6 EKSEFVCLGP 16

RESULT 8
Q8XY55 PRELIMINARY; PRT; 67 AA.
ID Q8XY55
AC Q8XY55
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Probable signal peptide protein.
GN RSC1908 OR RS03476.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoudat M., Gerin S., Artiguenave F., Gouzy J., Margenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Siglier P., Thebault P., Whalen M., Winkler P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
RL Nature 415:497-502(2002).
DR EMBL; AL646067; CAD15610.1; -.
KW Complete proteome.
SQ SEQUENCE 67 AA; 7293 MW; C64BDB4B64949B1 CRC64;

```

Query Match 59.7%; Score 40; DB 16; Length 67;  
Best Local Similarity 54.5%; Pred. No. 5;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 EKNSFECILGP 12  
DB 45 ESDTECVLDP 55

## RESULT 9

Q9YUD0 PRELIMINARY; PRT; 86 AA.

AC Q9YUD0; 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DE Hypothetical protein.  
OS Beak and feather disease virus.  
OC Viruses; ssDNA viruses; Circoviridae; Circovirus.  
OX NCBI\_Taxid=77856;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MELLINE=9909397; PubMed=9791035;  
RA Bassant M.R., Berryman D., Wilcox G.E., Raidal S.R.;  
RT "Palticache beak and feather disease virus nucleotide sequence analysis and its relationship to porcine circovirus, plant  
RT circoviruses, and chicken anaemia virus."  
RT Virology 249:453-459(1998).

RN [2]  
RP SEQUENCE FROM N.A.  
RA Bassant M.R., Berryman D., Wilcox G.E., Raidal S.R.;  
RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
RE EMBL: AF080560; AAC69867.1; -  
KM Hypothetical protein.  
KT VARIANT 51 N->T  
SQ SEQUENCE 86 AA; 8741 MW; AE1BB9AF83B09925 CRC64;

Query Match 59.7%; Score 40; DB 12; Length 86;  
Best Local Similarity 58.3%; Pred. No. 6.4;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 DEKNSFECILGP 12  
DB 19 DSKYLFECILAP 30

## RESULT 10

Q8NMNO PRELIMINARY; PRT; 236 AA.

AC Q8NMNO; 01-OCT-2002 (TREMBlrel. 22, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DE ABC-type transporter, ATPase component.  
GN CG12553.

OS Corynebacterium glutamicum (Brevibacterium flavum).  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
OX NCBI\_Taxid=1718;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;  
RA Nakagawa S.;

RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032,"  
RT Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
RE EMBL: AP005282; BAB99946.1; -

DR GO:0016020; C:membrane; IEA.  
DR GO:0005524; F:ATP binding; IEA.  
DR GO:0040009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
DR GO:0006810; P:transport; IEA.  
DR InterPro: IPR003439; ABC transporter.  
DR Pfam: PF00005; ABC\_tran\_1.

DR ProDom: PD000006; ABC transporter; 1.  
DR PROSITE: PS50893; ABC\_TRANSPORTER\_2; 1.  
KM Complete proteome.  
SQ SEQUENCE 236 AA; 25796 MW; 8738A5A2AD27BB3D CRC64;

Query Match 59.7%; Score 40; DB 16; Length 216;  
Best Local Similarity 63.6%; Pred. No. 17;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EKNSFECILGP 12  
DB 25 EKNSFTALNGP 35

## RESULT 11

Q8DVP3 PRELIMINARY; PRT; 289 AA.

AC Q8DVP3; 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DE Putative ABC transporter, ATP-binding protein.  
GN SMU\_431.

OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_Taxid=1309;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=UAI59 / ATCC 700610 / Serotype C;  
RX MEDLINE=22295063; PubMed=12397186;

RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,  
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,  
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;  
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental  
RT pathogen."  
RT Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).

RL EMBL: AE014889; AAN58183.1; -  
DR GO:0016020; C:membrane; IEA.  
DR GO:0005524; F:ATP binding; IEA.  
DR GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
DR GO:0000166; F:nucleotide binding; IEA.  
DR GO:0006810; P:transport; IEA.

DR InterPro: IPR003593; AAA\_ATPase.  
DR InterPro: IPR003439; ABC transporter.  
DR Pfam: PF00005; ABC\_tran\_1.  
DR ProDom: PD000006; ABC transporter; 1.  
DR SMART: SM00382; AAA; 1.  
DR PROSITE: PS00211; ABC\_TRANSPORTER\_1; 1.  
DR PROSITE: PS00893; ABC\_TRANSPORTER\_2; 1.  
ATP-binding; Complete proteome.

SQ SEQUENCE 289 AA; 32170 MW; C339DP9B86A5023 CRC64;

Query Match 59.7%; Score 40; DB 16; Length 289;  
Best Local Similarity 72.7%; Pred. No. 21;  
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 EKNSFECILGP 12  
DB 27 EKNSFTALNGP 37

## RESULT 12

Q87JL1 PRELIMINARY; PRT; 372 AA.

AC Q87JL1; 01-JUN-2003 (TREMBlrel. 24, Created)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DE ABC transporter, ATP-binding protein.  
GN VPA0238.

OS Vibrio parahaemolyticus.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;

```

OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIMD 2210633 / Serotype O3:k6;
RX MEDLINE=22508454; PubMed=12620739;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Najiima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shingawa H., Hattori M., Iida T.;
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT distinct from that of V. cholerae.";
RL Lancet 361:743-749 (2003).
DR EMBL: AF005084; BAC1581.1; -.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0005524; P:ATP binding; IEA.
DR GO: GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR003439; ABC_transporter.
DR Pfam: PF00005; ABC_tran; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE: PS50893; ABC_TRANSPORTER_2; 1.
DR ATP-binding; Complete proteome.
SQ SEQUENCE 372 AA; 4110 MW; 16BBFB07B9335863 CRC64;

Query Match 59.7%; Score 40; DB 16; Length 372;
Best Local Similarity 63.6%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 EKNSFECLIGP 12
DB 31 EKGEFVCFUGP 41

RESULT 13
OX KLY9
ID 09KLY9 PRELIMINARY; PRT; 403 AA.
AC 09KLY9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)
DE ABC transporter, ATP-binding protein.
GN VCA0602.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Drygiol I., Sellers P.,
RA McDonald L., Ueberlack T., Fleischmann R.D., Niernan W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483 (2000).
CC -1 SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
(ABC TRANSPORTERS).
DR EMBL: AE004390; AAP96503.1; -.
DR PIR: H82440; H82440.
DR HSSP: Q58663; IG6H.
DR TIGR: VCA0602; -.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0005524; P:ATP binding; IEA.
DR GO: GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR003593; AAA_ATPase.

```

```

DR InterPro: IPR003439; ABC_transporter.
DR Pfam: PF00005; ABC_tran; 1.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE: PS50893; ABC_TRANSPORTER_2; 1.
DR ATP-binding; Complete proteome.
SQ SEQUENCE 403 AA; 4456 MW; 867B28C76EA26272 CRC64;

Query Match 59.7%; Score 40; DB 16; Length 403;
Best Local Similarity 63.6%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 EKNSFECLIGP 12
DB 58 EKGEFVCFUGP 68

RESULT 14
OX UH40
ID 07UH40 PRELIMINARY; PRT; 455 AA.
AC 07UH40;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)
DE Hypothetical protein.
GN RA4864.
OS Rhodospirillum rubrum.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1;
RX MEDLINE=22735913; PubMed=12835416;
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heltmann K., Rabus R.,
RA Schlesner H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
RT strain 1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303 (2003).
DR EMBL: BX294141; CAD78139.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 455 AA; 49344 MW; 6EBA536F76ADEBD90 CRC64;

Query Match 59.0%; Score 39.5; DB 16; Length 455;
Best Local Similarity 66.7%; Pred. No. 40;
Matches 8; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

QY 1 DEKNSFECLIGP 11
DB 172 NEENSVEFCVIG 183

RESULT 15
OX 026593 PRELIMINARY; PRT; 187 AA.
AC 026593;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, last annotation update)
DE Hypothetical protein MTH493.
GN MTH493.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=96037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., Dubois J.,
RA Aldege T., Bashirzadeh R., Blakey D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pochier B., Qiu D.,

```

RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
 RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
 RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,  
 RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.,  
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
 RT deltaH: functional analysis and comparative genomics.",  
 RL J. Bacteriol. 179:7135-7155(1997).  
 DR EMBL; AE000833; AAB84999.1; -.  
 DR PIR; G69164; G69164.  
 KM Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 187 AA; 20693 MW; D778F23DFAC04032 CRC64;

Query March 58.2%; Score 39; DB 17; Length 187;  
 Best Local Similarity 54.5%; Pred. No. 21;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 DEKNSPECTIG 11  
 Db 28 DEHNMICVIG 38

Search completed: March 22, 2004, 06:59:16  
 Job time : 5.41467 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:30:23 ; Search time 3.64869 Seconds

(without alignments)  
929.256 Million cell updates/sec

Title: US-09-662-293-5

Perfect score: 67

Sequence: 1 DEMNSPFCILGP 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Tan04:\*

1: geneseqp1980s:\*

2: geneseqp1980s:\*

3: geneseqp2000s:\*

4: geneseqp2000s:\*

5: geneseqp2000s:\*

6: geneseqp2003as:\*

7: geneseqp2003bs:\*

8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	67	100.0	12	3	AAV52514	AAV52514 House dust
2	67	100.0	12	5	AAU96318	AAU96318 Der HMW-m
3	67	100.0	536	3	AAV52525	AAV52525 House dust
4	67	100.0	536	5	AAU96329	AAU96329 Der HMW-m
5	67	100.0	555	3	AAV52523	AAV52523 House dust
6	67	100.0	555	5	AAU96327	AAU96327 Der HMW-m
7	67	100.0	555	5	AAU96328	AAU96328 Der HMW-m
8	62	92.5	490	3	AAV52535	AAV52535 D. pteron
9	62	92.5	490	5	AAU96339	AAU96339 Der HMW-m
10	62	92.5	509	3	AAV52533	AAV52533 D. pteron
11	62	92.5	509	5	AAU96337	AAU96337 Der HMW-m
12	62	92.5	509	5	AAU96338	AAU96338 Der HMW-m
13	41	61.2	291	6	ABU48773	ABU48773 protein e
14	40	59.7	236	4	AAU92550	AAU92550 C. glutamini
15	40	59.7	284	4	AAU96334	AAU96334 Corynebacterium
16	40	59.7	284	4	AAU96332	AAU96332 Corynebacterium
17	40	59.7	1010	4	ABG14764	ABG14764 Novel human
18	40	59.7	1619	4	ABG18001	ABG18001 Novel human
19	39	58.2	99	4	ABG07930	ABG07930 Novel human
20	39	58.2	118	6	ADA34507	ADA34507 Actinobaculum
21	39	58.2	2703	4	ABE60266	ABE60266 Drosophila
22	38	56.7	51	4	AAU19287	AAU19287 peptide #
23	38	56.7	51	4	ABE38562	ABE38562 peptide #
24	38	56.7	51	4	AAU32012	AAU32012 peptide #
25	38	56.7	51	4	ABE23692	ABE23692 Protein #

26	38	56.7	51	4	AAW71720 Human bon
27	38	56.7	51	4	AAW59185 Human bra
28	38	56.7	51	4	ABG53405 Human liv
29	38	56.7	51	5	ABG41534 Human pep
30	38	56.7	155	7	ABE59316 Rat Prote
31	38	56.7	369	5	AAW51135 Common he
32	38	56.7	609	2	AAW71581 Human hep
33	38	56.7	659	7	ADC39096 Novel hum
34	38	56.7	686	6	AAE32625 Human pep
35	38	56.7	694	7	ADC49443 Human pro
36	38	56.7	1354	4	ABE60329 Drosophila
37	37.5	56.0	166	3	AAU96143 Arabidops
38	37.5	56.0	166	3	AAU96143 Arabidops
39	37.5	56.0	166	3	AAU96143 Arabidops
40	37.5	56.0	166	3	AAU96143 Arabidops
41	37.5	56.0	166	3	AAU96143 Arabidops
42	37.5	56.0	166	3	AAU96143 Arabidops
43	37.5	56.0	166	3	AAU96143 Arabidops
44	37.5	56.0	166	3	AAU96143 Arabidops
45	37.5	56.0	166	3	AAU96143 Arabidops

## ALIGNMENTS

AAV52514	RESULT 1	AAW71720
ID	AAV52514 standard; peptide; 12 AA.	AAW71720
XX	AAV52514; (first entry)	AAW59185
XX	22-FEB-2000	ABG53405
XX	House dust mite allergen protein (map) A/B fragment map(4).	ABG41534
XX	Mite allergen protein; map; high molecular weight; HMW-map; allergy;	ABE59316
XX	house dust mite; IGE; immunoglobulin E; allergen; mapA; mapB;	AAW51135
XX	hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;	AAW71581
XX	canine; veterinary; antibody; vaccine; immunisation.	ADC39096
OS	Dermatophagoides farinae.	AAE32625
XX	W09954349-A2.	ADC49443
XX	28-OCT-1999.	ABE60329
XX	16-APR-1999; 99WO-US008524.	AAU96143
XX	17-APR-1998; 98US-00062013.	AAU96143
XX	13-MAY-1998; 98US-0085295P.	AAU96143
XX	02-SEP-1998; 98US-0098909P.	AAU96143
XX	(HESK-) HESKA CORP.	AAU96143
XX	Mccall CA, Hunter SM, Weber ER;	AAU96143
XX	WPI, 2000-052700/04.	AAU96143
XX	Novel high molecular weight Dermatophagoides nucleic acid polypeptides	AAU96143
XX	used to modify an animals' hypersensitivity to mite allergens.	AAU96143
XX	Claim 3; Page 69; 154pp; English.	AAU96143
XX	Sequences AAV52510-Y52522 represent proteolytic fragments of	AAU96143
XX	Dermatophagoides farinae high molecular weight mite allergen protein (HMW	AAU96143
XX	-map) composition. The HMW-map composition was isolated from a D. farinae	AAU96143
XX	homogenate by gel filtration, with each fraction being analysed for the	AAU96143
XX	presence of proteins that bound to IGE present in mite-allergic dog	AAU96143
XX	antisera. The HMW-map composition comprises mapA (a 109 kD protein) and	AAU96143
XX	mapB (98 kD). Mite allergenic proteins and peptides, and nucleic acids	AAU96143
XX	encoding them, may be used in therapeutic compositions to modify an	AAU96143
XX	animal's hypersensitivity reaction to mite allergens. Animals that may be	AAU96143
XX	treated include mammals and birds, especially felines, canines, equines,	AAU96143

CC humans, other pets, and work or domestic animals. The proteins or  
 CC fragments may also be used to diagnose allergies via a skin test. The  
 CC proteins and peptides can also be used to raise antibodies, which have a  
 CC variety of potential uses. For example, they can be used as vaccines to  
 CC passively immunise animals against dust mite hypersensitivity, as  
 CC positive controls in test kits and as tools to recover desired dust mite  
 CC allergens from a mixture of proteins

XX Sequence 12 AA;

Query Match 100.0%; Score 67; DB 3; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 4.8e-05;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DEKNSPECTILGP 12  
 DB 1 DEKNSPECTILGP 12

RESULT 2  
 AAU96318  
 ID AAU96318 standard; peptide; 12 AA.

XX AAU96318;  
 AC  
 XX 15-JUL-2002 (first entry)  
 DT  
 XX  
 DE Der HMW-map polypeptide #5.  
 XX

KM Der HMW-map; American house dust mite; antiallergic; mite; IgE;  
 KM mite allergenic protein; immunoglobulin E; hypersensitivity;  
 KM immunocomplex formation.  
 XX

OS Dermatophagoides farinae.

XX WO200222807-A2.

XX 21-MAR-2002.

XX 14-SEP-2001; 2001WO-US028730.

XX 14-SEP-2000; 2000US-00662293.

XX (HESK-) HESKA CORP.

XX Mccall CA, Hunter SW, Weber ER;

XX WPI; 2002-351888/38.

PT New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HMW-map protein, useful as a vaccine for treating mite allergy.

XX Claim 12; Page 70; 16pp; English.

CC The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic  
 CC acid. The Der HMW-map protein is useful for eliciting an immune response  
 CC against Der HMW-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitising a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig) E or Der HMW-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HMW-map polypeptides of the invention

XX Sequence 12 AA;

Query Match 100.0%; Score 67; DB 5; Length 12;

Best Local Similarity 100.0%; Pred. No. 4.8e-05;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DEKNSPECTILGP 12  
 DB 1 DEKNSPECTILGP 12

RESULT 3  
 AAU52525

ID AAU52525 standard; protein; 536 AA.

AC AAU52525;

DT 22-FEB-2000 (first entry)

DE House dust mite (D. farinae) mite allergen protein (map) Pderf98-536.

KM Mite allergen protein; map; high molecular weight; HMW-map; allergy;  
 KM house dust mite; IgE; immunoglobulin E; allergen; mapA; mapB;  
 KM hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
 KM canine; veterinary; antibody; vaccine; immunisation.  
 XX

OS Dermatophagoides farinae.

XX WO954349-A2.

XX 28-OCT-1999.

XX 16-APR-1999; 99WO-US008524.

XX 17-APR-1998; 98US-00062013.

XX 13-MAY-1998; 98US-0085295P.

XX 02-SEP-1998; 98US-0098909P.

XX (HESK-) HESKA CORP.

XX Mccall CA, Hunter SW, Weber ER;

XX WPI; 2000-052700/04.

XX N-PSDB; AA238579, AA238580.

PT Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
 PT used to modify an animals' hypersensitivity to mite allergens.

XX Claim 3; Page 125-127; 154pp; English.

CC This sequence represents Dermatophagoides farinae mite allergen protein  
 CC (map) Pderf98-536, the mature form of Pderf98-555 (AAU52523). Pderf98-536  
 CC has a molecular weight of 98 kD, comprising 536 amino acids, and is a  
 CC component of the Dermatophagoides farinae high molecular weight mite  
 CC allergen protein (HMW-map) composition. The HMW-map composition was  
 CC isolated from a D. farinae homogenate by gel filtration, with each  
 CC fraction being analysed for the presence of proteins that bound to IgE  
 CC present in mite-allergic dog antisera. Mite allergenic proteins and  
 CC peptides, and nucleic acids encoding them, may be used in therapeutic  
 CC compositions to modify an animal's hypersensitivity reaction to mite  
 CC allergens. Animals that may be treated include mammals and birds,  
 CC especially felines, canines, equines, humans, other pets, and work or  
 CC domestic animals. The proteins or fragments may also be used to diagnose  
 CC allergies via a skin test. The proteins and peptides can also be used to  
 CC raise antibodies, which have a variety of potential uses. For example,  
 CC they can be used as vaccines to passively immunise animals against dust  
 CC mite hypersensitivity, as positive controls in test kits and as tools to  
 CC recover desired dust mite allergens from a mixture of proteins

XX Sequence 536 AA;

Query Match 100.0%; Score 67; DB 3; Length 536;  
 Best Local Similarity 100.0%; Pred. No. 0.0021;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DEKNSPECTILGP 12

Db 385 DEKNSFECTILGP 396

RESULT 4  
AAU96329  
ID AAU96329 standard; protein; 536 AA.  
XX  
XX AAU96329;  
AC  
XX 15-JUL-2002 (first entry)  
DT  
XX  
DE Der HMW-map polypeptide #16.  
XX  
XX Der HMW-map; American house dust mite; antiallergic; mite; IGE;  
KM mite allergenic protein; immunoglobulin E; hypersensitivity;  
XX immunocomplex formation.  
XX  
XX Dermatophagoides farinae.  
OS  
XX WO200222807-A2.  
PN  
XX 21-MAR-2002.  
PD  
XX 14-SEP-2001; 2001WO-US028730.  
PF  
XX 14-SEP-2000; 2000US-00662293.  
PR  
XX (HESK-) HESKA CORP.  
PA  
XX Mccall CA, Hunter SW, Weber ER;  
PI  
XX WPI; 2002-351888/38.  
DR N-PSDB; ABK69575.  
XX  
XX New mite allergenic protein isolated from Dermatophagoides, designated  
PT Der HMW-map protein, useful as a vaccine for treating mite allergy.  
PT  
XX  
PS Claim 12; Page 125-127; 16pp; English.  
XX  
XX The invention relates to an isolated mite allergenic protein of  
CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic  
CC acid. The Der HMW-map protein is useful for eliciting an immune response  
CC against Der HMW-map protein. The protein or a reagent comprising a non-  
CC proinfectious epitope is useful for identifying an animal (e.g., dog,  
CC cat) susceptible to or having an allergic response to a mite. A  
CC therapeutic composition is useful for desensitizing a host animal to an  
CC allergic response to a mite. The DNA and protein can be used in the  
CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
CC of immunoglobulin (Ig) E or Der HMW-map protein activity associated with a  
CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
CC binding of proteins to IGE, to prevent immunocomplex formation, thus  
CC reducing hypersensitivity responses to mite allergens, and as vaccines  
CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
CC represent Der HMW-map polypeptides of the invention  
XX  
XX  
SQ Sequence 536 AA;

Query Match 100.0%; Score 67; DB 5; Length 536;  
Best Local Similarity 100.0%; Pred. No. 0.0021;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DEKNSFECTILGP 12  
Db 385 DEKNSFECTILGP 396

RESULT 5  
AAU96323  
ID AAU96323 standard; protein; 555 AA.  
XX  
XX AAU96323;  
AC  
XX

DT 22-FEB-2000 (first entry)  
XX  
XX House dust mite (D. farinae) mite allergen protein (map) Pderf98-555.  
DE  
XX Mite allergen protein; map; high molecular weight; HMW-map; allergy;  
XX house dust mite; IGE; immunoglobulin E; allergen; mapA; mapB;  
KM hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
KM canine; veterinary; antibody; vaccine; immunisation.  
XX  
XX Dermatophagoides farinae.  
OS  
XX  
XX  
XX Key Location/Qualifiers  
FT Peptide 1..19  
FT /note= "Signal peptide"  
FT Protein 20..555  
FT /note= "Mature Pderf98-555"

XX WO9954349-A2.  
XX  
XX 28-OCT-1999.  
PD  
XX 16-APR-1999; 99WO-US008524.  
XX  
XX 17-APR-1998; 98US-00062013.  
XX 13-MAY-1998; 98US-0085235P.  
PR 02-SEP-1998; 98US-0098909P.  
XX  
XX (HESK-) HESKA CORP.  
PA  
XX Mccall CA, Hunter SW, Weber ER;  
PI  
XX WPI; 2000-052700/04.  
DR N-PSDB; AAZ38575; AAZ38576; AAZ38577; AAZ38578.  
XX  
XX Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
PT used to modify an animal's hypersensitivity to mite allergens.  
PT  
XX  
PS Claim 3; Page 111-113; 154pp; English.

XX This sequence represents Dermatophagoides farinae mite allergen protein  
CC (map) Pderf98-555. Pderf98-555 has a molecular weight of 98 kD,  
CC comprising 555 amino acids, and is a component of the Dermatophagoides  
CC farinae high molecular weight mite allergen protein (HMW-map)  
CC composition. The HMW-map composition was isolated from a D. farinae  
CC homogenate by gel filtration, with each fraction being analysed for the  
CC presence of proteins that bound to IGE present in mite-allergic dog  
CC antisera. Mite allergenic proteins and peptides, and nucleic acids  
CC encoding them, may be used in therapeutic compositions to modify an  
CC animal's hypersensitivity reaction to mite allergens. Animals that may be  
CC treated include mammals and birds, especially felines, canines, equines,  
CC humans, other pets, and work or domestic animals. The proteins or  
CC fragments may also be used to diagnose allergies via a skin test. The  
CC proteins and peptides can also be used to raise antibodies, which have a  
CC variety of potential uses. For example, they can be used as vaccines to  
CC passively immunise animals against dust mite hypersensitivity, as  
CC positive controls in test kits and as tools to recover desired dust mite  
CC allergens from a mixture of proteins  
XX  
XX  
SQ Sequence 555 AA;

Query Match 100.0%; Score 67; DB 3; Length 555;  
Best Local Similarity 100.0%; Pred. No. 0.0021;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DEKNSFECTILGP 12  
Db 404 DEKNSFECTILGP 415

RESULT 6  
AAU96327  
ID AAU96327 standard; protein; 555 AA.  
XX  
XX

AC AAU96327;  
 XX 15-JUL-2002 (first entry)  
 XX  
 DE Der HMW-map polypeptide #14.  
 XX  
 XX Der HMW-map; American house dust mite; anti-allergic; mite; IgE;  
 KM mite allergenic protein; immunoglobulin E; hypersensitivity;  
 KM immunocomplex formation.  
 OS Dermatophagoides farinae.  
 XX WO200222807-A2.  
 XX  
 XX 21-MAR-2002.  
 XX  
 XX 14-SEP-2001; 2001WO-US028730.  
 XX  
 XX 14-SEP-2000; 2000US-00662293.  
 XX  
 XX (HESK-) HESKA CORP.  
 PA  
 PI Mccall CA, Hunter SW, Weber ER;  
 XX  
 XX WPI; 2002-351888/38.  
 DR N-PSDB; ABK69571.  
 XX  
 XX New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HMW-map protein, useful as a vaccine for treating mite allergy.  
 XX  
 XX Claim 12; Page 114-116; 161pp; English.  
 PS  
 CC The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic  
 CC acid. The Der HMW-map protein is useful for eliciting an immune response  
 CC against Der HMW-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitising a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig) E or Der HMW-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HMW-map polypeptides of the invention  
 CC  
 XX  
 SQ Sequence 555 AA;

Query Match 100.0%; Score 67; DB 5; Length 555;  
 Best Local Similarity 100.0%; Pred. No. 0.0022;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DEKNSPECTIGP 12  
 DB 404 DEKNSPECTIGP 415

RESULT 7  
 AAU96328  
 ID AAU96328 standard; protein; 555 AA.  
 XX  
 AC AAU96328;  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX  
 DE Der HMW-map polypeptide #15.  
 XX  
 XX Der HMW-map; American house dust mite; anti-allergic; mite; IgE;  
 KM mite allergenic protein; immunoglobulin E; hypersensitivity;  
 KM immunocomplex formation.  
 XX

OS Dermatophagoides farinae.  
 XX  
 XX WO200222807-A2.  
 XX  
 PD 21-MAR-2002.  
 XX  
 XX 14-SEP-2001; 2001WO-US028730.  
 XX  
 XX 14-SEP-2000; 2000US-00662293.  
 XX  
 XX (HESK-) HESKA CORP.  
 PA  
 PI Mccall CA, Hunter SW, Weber ER;  
 XX  
 XX WPI; 2002-351888/38.  
 DR N-PSDB; ABK69573.  
 XX  
 XX New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HMW-map protein, useful as a vaccine for treating mite allergy.  
 XX  
 XX Claim 12; Page 120-122; 161pp; English.  
 PS  
 CC The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic  
 CC acid. The Der HMW-map protein is useful for eliciting an immune response  
 CC against Der HMW-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitising a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig) E or Der HMW-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HMW-map polypeptides of the invention  
 CC  
 XX  
 SQ Sequence 555 AA;

Query Match 100.0%; Score 67; DB 5; Length 555;  
 Best Local Similarity 100.0%; Pred. No. 0.0022;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DEKNSPECTIGP 12  
 DB 404 DEKNSPECTIGP 415

RESULT 8  
 AAU96335  
 ID AAU96335 standard; protein; 490 AA.  
 XX

AC AAU96335;  
 XX  
 DT 06-AUG-2003 (revised)  
 DT 22-FEB-2000 (first entry)  
 XX

DE D. pteronyssinus 98 kD mite allergen protein (map) PDerp98-490.  
 XX  
 XX Mite allergen protein; map; high molecular weight; HMW-map; allergy;  
 KM house dust mite; IgE; immunoglobulin E; allergen; map;  
 KM hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
 KM canine; veterinary; antibody; vaccine; immunisation.  
 XX  
 OS Dermatophagoides pteronyssinus.  
 XX  
 FH Key Location/Qualifiers  
 FH Modified-site 115..117  
 FT /note="Asn is N-glycosylated"  
 FT Modified-site 240..242  
 FT /note="Asn is N-glycosylated"  
 XX



```

PN WO9954349-A2.
XX
XX 28-OCT-1999.
XX
XX PF 16-APR-1999; 99WO-US008524.
XX
XX PR 17-APR-1998; 98US-00062013.
XX PR 13-MAY-1998; 98US-0085295P.
XX PR 02-SEP-1998; 98US-0086909P.
XX
XX PA (HESK-) HESKA CORP.
XX
XX PI Mccall CA, Hunter SW, Weber ER;
XX
XX DR WPI: 2000-052700/04.
XX DR N-PSDB; AA238589, AA238590.
XX
XX PT Novel high molecular weight Dermatophagoides nucleic acid polypeptides
XX PT used to modify an animals' hypersensitivity to mite allergens.
XX
XX PS Claim 3; Page 147-149; 154pp; English.
XX
XX CC This sequence represents Dermatophagoides pteronyssinus mite allergen
XX CC protein (map) Pderp98-490, the mature form of Pderp98-509. Pderp98-490
XX CC has a molecular weight of 98 kD, comprising 490 amino acids, and has a
XX CC high degree of homology with the D. farinae mature 98 kD allergen, mapB
XX CC (AA952525). Nucleic acid molecules encoding Pderp98-490 were isolated
XX CC from a D. pteronyssinus cDNA library by hybridisation with a probe
XX CC encoding the D. farinae high molecular weight map (Hmw-map) composition.
XX CC Mite allergenic proteins and peptides, and nucleic acids encoding them,
XX CC may be used in therapeutic compositions to modify an animals'
XX CC hypersensitivity reaction to mite allergens. Animals that may be treated
XX CC include mammals and birds, especially felines, canines, equines, humans,
XX CC other pets, and work or domestic animals. The proteins or fragments may
XX CC also be used to diagnose allergies via a skin test. The proteins and
XX CC peptides can also be used to raise antibodies, which have a variety of
XX CC potential uses. For example, they can be used as vaccines to passively
XX CC immunise animals against dust mite hypersensitivity, as positive controls
XX CC in test kits and as tools to recover desired dust mite allergens from a
XX CC mixture of proteins. (Updated on 06-AUG-2003 to correct OS field.)
XX
XX SQ Sequence 490 AA;

```

```

Query Match          92.5%; Score 62; DB 3; Length 490;
Best Local Similarity 83.3%; Pred. No. 0.015;
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 DEKNSPECTLGP 12
   |||||:||||
Db 385 DEKNSYECLGP 396

```

```

RESULT 9
AAU96339
ID AAU96339 standard; protein; 490 AA.
XX
XX AC AAU96339;
XX
XX DT 15-JUL-2002 (first entry)
XX
XX DE Der Hmw-map polypeptide #26.
XX
XX KW Der Hmw-map; American house dust mite; anti-allergic; mite; IGE;
XX KW mite allergenic protein; immunoglobulin E; hypersensitivity;
XX KW immunocomplex formation.
XX
XX OS Dermatophagoides farinae.
XX
XX PN WO200222807-A2.
XX
XX PD 21-MAR-2002.
XX
XX PF 14-SEP-2001; 2001WO-US028730.

```

```

XX
XX PR 14-SEP-2000; 2000US-00662293.
XX
XX PA (HESK-) HESKA CORP.
XX
XX PI Mccall CA, Hunter SW, Weber ER;
XX
XX DR WPI: 2002-351888/38.
XX DR N-PSDB; ABK69585.
XX
XX PT New mite allergenic protein isolated from Dermatophagoides, designated
XX PT Der Hmw-map protein, useful as a vaccine for treating mite allergy.
XX
XX PS Claim 12; Page 144-146; 161pp; English.
XX
XX CC The invention relates to an isolated mite allergenic protein of
XX CC Dermatophagoides, designated Der Hmw-map protein, and its related nucleic
XX CC acid. The Der Hmw-map protein is useful for eliciting an immune response
XX CC against Der Hmw-map protein. The protein or a reagent comprising a non-
XX CC proteinaceous epitope is useful for identifying an animal (e.g., dog,
XX CC cat) susceptible to or having an allergic response to a mite. A
XX CC therapeutic composition is useful for desensitising a host animal to an
XX CC allergic response to a mite. The DNA and protein can be used in the
XX CC detection of anti-Der Hmw-map antibodies in animal fluids, and inhibition
XX CC of immunoglobulin (Ig)E or Der Hmw-map protein activity associated with a
XX CC disease. Antibodies that bind to Der Hmw-map are useful for inhibiting
XX CC binding of proteins to IgE, to prevent immunocomplex formation, thus
XX CC reducing hypersensitivity responses to mite allergens, and as vaccines
XX CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342
XX CC represent Der Hmw-map polypeptides of the invention
XX
XX SQ Sequence 490 AA;

```

```

Query Match          92.5%; Score 62; DB 3; Length 490;
Best Local Similarity 83.3%; Pred. No. 0.015;
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 DEKNSPECTLGP 12
   |||||:||||
Db 385 DEKNSYECLGP 396

```

```

RESULT 10
AA52533
ID AA52533 standard; protein; 509 AA.
XX
XX AC AA52533;
XX
XX DT 06-AUG-2003 (revised)
XX DT 22-FEB-2000 (first entry)
XX
XX DE D. pteronyssinus 98 kD mite allergen protein (map) Pderp98-509.
XX
XX KW Mite allergen protein; map; high molecular weight; Hmw-map; allergy;
XX KW house dust mite; IGE; immunoglobulin E; allergen; mapB;
XX KW hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;
XX KW canine; veterinary; antibody; vaccine; immunisation.
XX
XX OS Dermatophagoides pteronyssinus.
XX
XX FH Key
XX FH Location/Qualifiers
XX FT Peptide 1..19
XX FT Protein /note= "Signal peptide"
XX FT /note= "Mature Pderp98-509"
XX
XX WO9954349-A2.
XX
XX PD 28-OCT-1999.
XX
XX PF 16-APR-1999; 99WO-US008524.
XX
XX PR 17-APR-1998; 98US-00062013.

```

PR 13-MAY-1998; 98US-0085295P.  
 PR 02-SEP-1998; 98US-0098909P.  
 XX  
 PA (HESK-) HESKA CORP.  
 XX  
 PI McCall CA, Hunter SW, Weber ER;  
 XX  
 DR WPI; 2000-052700/04.  
 DR N-PSDB; AA238585, AA238586, AA238587, AA238588.  
 XX  
 PT Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
 PT used to modify an animal's hypersensitivity to mite allergens.  
 XX  
 PS Claim 3; Page 134-136; 154pp; English.  
 XX  
 CC This sequence represents Dermatophagoides pteronyssinus mite allergen  
 CC protein (map) pterp98-509. Pterp98-509 has a molecular weight of 98 kD,  
 CC comprising 509 amino acids, and has a high degree of homology with the D.  
 CC farinae 98 kD allergen, map8 (AAV52523). Nucleic acid molecules encoding  
 CC pterp98-509 were isolated from a D. pteronyssinus cDNA library by  
 CC hybridisation with a probe encoding the D. farinae high molecular weight  
 CC map (HMM-map) composition. Mite allergenic proteins and peptides, and  
 CC nucleic acids encoding them, may be used in therapeutic compositions to  
 CC modify an animal's hypersensitivity reaction to mite allergens. Animals  
 CC that may be treated include mammals and birds, especially felines.  
 CC canines, equines, humans, other pets, and work or domestic animals. The  
 CC proteins or fragments may also be used to diagnose allergies via a skin  
 CC test. The proteins and peptides can also be used to raise antibodies,  
 CC which have a variety of potential uses. For example, they can be used as  
 CC vaccines to passively immunise animals against dust mite  
 CC hypersensitivity, as positive controls in test kits and as tools to  
 CC recover desired dust mite allergens from a mixture of proteins. (Updated  
 CC on 06-AUG-2003 to correct OS field.)  
 CC  
 XX  
 SQ Sequence 509 AA;  
 Query Match 92.5%; Score 62; DB 3; Length 509;  
 Best Local Similarity 83.3%; Pred. NO. 0.015; Indels 0; Gaps 0;  
 Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DEKNSFECTILGP 12  
 |||||:||||  
 Db 404 DEKNSYECULGP 415  
 RESULT 11  
 AAU96337  
 ID AAU96337 standard; protein; 509 AA.  
 XX  
 AC AAU96337;  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX  
 DE Der HMM-map polypeptide #24.  
 XX  
 KW Der HMM-map; American house dust mite; antiallergic; mite; IGE;  
 KW mite allergenic protein; immunoglobulin E; hypersensitivity;  
 KW immunocomplex formation.  
 XX  
 OS Dermatophagoides farinae.  
 XX  
 WO200222807-A2.  
 FN  
 PI McCall CA, Hunter SW, Weber ER;  
 XX  
 DR WPI; 2002-351888/38.  
 DR N-PSDB; ABK69581.  
 XX  
 PT New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HMM-map protein, useful as a vaccine for treating mite allergy.  
 XX  
 PS Claim 12; Page 134-136; 161pp; English.  
 XX  
 CC The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HMM-map protein, and its related nucleic  
 CC acid. The Der HMM-map protein is useful for eliciting an immune response  
 CC against Der HMM-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitising a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMM-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig) E or Der HMM-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HMM-map are useful for inhibiting  
 CC binding of proteins to IGE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HMM-map polypeptides of the invention  
 CC  
 XX  
 SQ Sequence 509 AA;  
 Query Match 92.5%; Score 62; DB 5; Length 509;  
 Best Local Similarity 83.3%; Pred. NO. 0.015; Indels 0; Gaps 0;  
 Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DEKNSFECTILGP 12  
 |||||:||||  
 Db 404 DEKNSYECULGP 415

DR WPI; 2002-351888/38.  
 DR N-PSDB; ABK69581.  
 XX  
 PT New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HMM-map protein, useful as a vaccine for treating mite allergy.  
 XX  
 PS Claim 12; Page 134-136; 161pp; English.  
 XX  
 CC The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HMM-map protein, and its related nucleic  
 CC acid. The Der HMM-map protein is useful for eliciting an immune response  
 CC against Der HMM-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitising a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMM-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig) E or Der HMM-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HMM-map are useful for inhibiting  
 CC binding of proteins to IGE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HMM-map polypeptides of the invention  
 CC  
 XX  
 SQ Sequence 509 AA;  
 Query Match 92.5%; Score 62; DB 5; Length 509;  
 Best Local Similarity 83.3%; Pred. NO. 0.015; Indels 0; Gaps 0;  
 Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DEKNSFECTILGP 12  
 |||||:||||  
 Db 404 DEKNSYECULGP 415  
 RESULT 12  
 AAU96338  
 ID AAU96338 standard; protein; 509 AA.  
 XX  
 AC AAU96338;  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX  
 DE Der HMM-map polypeptide #25.  
 XX  
 KW Der HMM-map; American house dust mite; antiallergic; mite; IGE;  
 KW mite allergenic protein; immunoglobulin E; hypersensitivity;  
 KW immunocomplex formation.  
 XX  
 OS Dermatophagoides farinae.  
 XX  
 WO200222807-A2.  
 FN  
 PI McCall CA, Hunter SW, Weber ER;  
 XX  
 DR WPI; 2002-351888/38.  
 DR N-PSDB; ABK69583.  
 XX  
 PT New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HMM-map protein, useful as a vaccine for treating mite allergy.  
 XX  
 PS Claim 12; Page 139-141; 161pp; English.  
 XX  
 CC The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HMM-map protein, and its related nucleic

CC against. The Der HMW-map protein is useful for eliciting an immune response  
CC against Der HMW-map protein. The protein or a reagent comprising a non-  
CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
CC cat) susceptible to or having an allergic response to a mite. A  
CC therapeutic composition is useful for desensitizing a host animal to an  
CC allergic response to a mite. The DNA and protein can be used in the  
CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
CC of immunoglobulin (IgE or Der HMW-map protein activity associated with a  
CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
CC reducing hypersensitivity responses to mite allergens, and as vaccines  
CC against mite allergen hypersensitivity. Sequences AAU6314-AAU6342  
CC represent Der HMW-map polypeptides of the invention  
CC  
SQ Sequence 509 AA;  
SQ  
Query Match 92.5%; Score 62; DB 5; Length 509;  
Best Local Similarity 83.3%; Pred. No. 0.015;  
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DEKNSPECTILGP 12  
|||||:|||||  
Db 404 DEKNSYECTILGP 415  
Db  
RESULT 13  
ABU48773  
ID ABU48773 standard; protein; 291 AA.  
XX  
AC ABU48773;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #34300.  
XX  
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
OS Ureaplasma urealyticum.  
XX  
PN WC000277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WC-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
XX  
PR 06-SEP-2001; 2001US-00948993.  
XX  
PR 25-OCT-2001; 2001US-0342923P.  
XX  
PR 08-FEB-2002; 2002US-00072851.  
XX  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KJ, Zyskind JW;  
PI Walli D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR MPI; 2003-029926/02.  
XX  
DR N-PSDB; ACAS2643.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids, required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 76697; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pcr\_sequences  
XX  
SQ Sequence 291 AA;  
SQ  
Query Match 61.2%; Score 41; DB 6; Length 291;  
Best Local Similarity 72.7%; Pred. No. 50;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 2 EKNSPECTILGP 12  
:|||||  
Db 53 KONSFTTILGP 63  
Db  
RESULT 14  
AAG92550  
ID AAG92550 standard; protein; 236 AA.  
XX  
AC AAG92550;  
XX  
DT 26-SEP-2001 (first entry)  
XX  
DE C glutamicum protein fragment SEQ ID NO: 6304.  
XX  
KW Corynebacterium; amino acid synthesis; vitamin; saccharide;  
XX  
KW organic acid synthesis.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN EP1108790-A2.  
XX  
FN  
XX  
PD 20-JUN-2001.  
XX  
PF 18-DEC-2000; 2000EP-00127668.  
XX  
PR 16-DEC-1999; 99JP-00377484.  
XX  
PR 07-APR-2000; 2000JP-00159162.  
XX  
PR 03-AUG-2000; 2000JP-00280988.  
XX  
PA (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI Tachishi N, Senoh A, Ikeda M, Ozaki A;  
XX  
DR MPI; 2001-376931/40.  
XX  
DR N-PSDB; AAH67769.  
XX  
PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analyzing  
PT expression profile or pattern of a gene and identifying homologous gene.  
XX  
PS Claim 17; SEQ ID NO 6304; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein sequences from the Corynebacterium glutamicum. These are useful for identifying the mutation point of a gene derived from a mutant of Corynebacterium glutamicum, measuring expression amount and analyzing the expression profile or expression pattern of a gene derived from Corynebacterium glutamicum, and identifying a homologue of a gene derived from Corynebacterium glutamicum. Corynebacterium glutamicum are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a protein described in the exemplification of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office

XX Sequence 236 AA;

Query Match 59.7%; Score 40; DB 4; Length 236;  
Best Local Similarity 63.6%; Pred. No. 61;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EKNSFECILGP 12  
DB 25 EKNKFTAIMGP 35

RESULT 15  
AAB76734 ID AAB76734 standard; protein; 284 AA.

XX AAB76734;  
XX 11-APR-2001 (first entry)  
DE Corynebacterium glutamicum MCT protein SEQ ID NO:450.  
XX  
XX Corynebacterium glutamicum; Brevibacterium lactofermentum; MCT; membrane construction and membrane transport protein; petroleum spill; hydrocarbon degradation; gram positive aerobic bacterium; marker; identification; microorganism; fine chemical production; transformation; genome mapping; genetic engineering.  
XX  
XX Corynebacterium glutamicum.  
XX  
XX MO200100805-A2.  
XX  
XX 04-JAN-2001.  
XX  
XX 23-JUN-2000; 2000WO-IB000926.  
XX  
XX 25-JUN-1999; 99US-0141031P.  
PR 08-JUL-1999; 99DE-01031454.  
PR 08-JUL-1999; 99DE-01031478.  
PR 08-JUL-1999; 99DE-01031563.  
PR 09-JUL-1999; 99DE-01032122.  
PR 09-JUL-1999; 99DE-01032124.  
PR 09-JUL-1999; 99DE-01032125.  
PR 09-JUL-1999; 99DE-01032128.  
PR 09-JUL-1999; 99DE-01032180.  
PR 09-JUL-1999; 99DE-01032182.  
PR 09-JUL-1999; 99DE-01032190.  
PR 09-JUL-1999; 99DE-01032191.  
PR 09-JUL-1999; 99DE-01032209.  
PR 09-JUL-1999; 99DE-01032212.  
PR 09-JUL-1999; 99DE-01032227.  
PR 09-JUL-1999; 99DE-01032228.  
PR 09-JUL-1999; 99DE-01032229.  
PR 09-JUL-1999; 99DE-01032230.  
PR 14-JUL-1999; 99DE-01032927.  
PR 14-JUL-1999; 99DE-01033005.  
PR 14-JUL-1999; 99DE-01033006.  
PR 27-AUG-1999; 99DE-01040764.  
PR 27-AUG-1999; 99DE-01040765.  
PR 27-AUG-1999; 99DE-01040766.

PR 27-AUG-1999; 99DE-01040830.  
PR 27-AUG-1999; 99DE-01040831.  
PR 27-AUG-1999; 99DE-01040832.  
PR 27-AUG-1999; 99DE-01040833.  
PR 31-AUG-1999; 99DE-01041378.  
PR 31-AUG-1999; 99DE-01041379.  
PR 31-AUG-1999; 99DE-01041395.  
PR 03-SEP-1999; 99DE-01042077.  
PR 03-SEP-1999; 99DE-01042078.  
PR 03-SEP-1999; 99DE-01042079.  
PR 03-SEP-1999; 99DE-01042088.  
XX  
XX (BADI ) BASF AG.  
XX  
XX Pompeius M, Kroeger B, Schroeder H, Zelder O, Habermayer G;  
XX  
XX MPI; 2001-071486/08.  
XX  
XX N-PSDB; AAF67967.  
XX  
XX Corynebacterium glutamicum nucleic acids encoding membrane construction and membrane transport proteins or their portions, useful for typing or identifying C. glutamicum or related bacteria, and as markers for transformation.  
XX  
XX Claim 20; Page 796-797; 1119pp; English.  
XX  
XX AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane construction and membrane transport (MCT) proteins given in AAB76510 to AAB76847. The MCT nucleic acids and proteins are useful in the identification of microorganisms which can be used to produce fine chemicals, for modulating fine chemical production in C. glutamicum or related bacteria (e.g. Brevibacterium lactofermentum), the typing or identification of C. glutamicum or related bacteria, as reference points for mapping C. glutamicum genome, and as markers for transformation.  
XX  
XX AAF68082 and AAF68082 represent sequencing primers which are used in an example from the present invention  
XX  
XX Sequence 284 AA;

QY 2 EKNSFECILGP 12  
DB 73 EKNKFTAIMGP 83

Query Match 59.7%; Score 40; DB 4; Length 284;  
Best Local Similarity 63.6%; Pred. No. 74;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Search completed: March 22, 2004, 06:51:38  
Job time : 5.64869 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: March 22, 2004, 06:59:34 ; Search time 2.4547 Seconds

(without alignments)  
1265.926 Million cell updates/sec

Title: US-09-662-293-5

Perfect score: 1 DEKNSPECTIGP 12

Sequence: BLOSUM62

Scoring table: Gapop 10.0 , Gapext 0.5

Searched: 1049977 seqs, 258955339 residues

Total number of hits satisfying chosen parameters: 1049977

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

## Database:

```
1: /cgml_6/ptodata/2/pubppa/US07_PUBCOMB.pep:*
2: /cgml_6/ptodata/2/pubppa/US07_PUBCOMB.pep:*
3: /cgml_6/ptodata/2/pubppa/US06_PUBCOMB.pep:*
4: /cgml_6/ptodata/2/pubppa/US06_PUBCOMB.pep:*
5: /cgml_6/ptodata/2/pubppa/US07_PUBCOMB.pep:*
6: /cgml_6/ptodata/2/pubppa/US07_PUBCOMB.pep:*
7: /cgml_6/ptodata/2/pubppa/US07_PUBCOMB.pep:*
8: /cgml_6/ptodata/2/pubppa/US08_PUBCOMB.pep:*
9: /cgml_6/ptodata/2/pubppa/US08_PUBCOMB.pep:*
10: /cgml_6/ptodata/2/pubppa/US08_PUBCOMB.pep:*
11: /cgml_6/ptodata/2/pubppa/US09_PUBCOMB.pep:*
12: /cgml_6/ptodata/2/pubppa/US09_PUBCOMB.pep:*
13: /cgml_6/ptodata/2/pubppa/US10_PUBCOMB.pep:*
14: /cgml_6/ptodata/2/pubppa/US10_PUBCOMB.pep:*
15: /cgml_6/ptodata/2/pubppa/US10_PUBCOMB.pep:*
16: /cgml_6/ptodata/2/pubppa/US10_PUBCOMB.pep:*
17: /cgml_6/ptodata/2/pubppa/US10_PUBCOMB.pep:*
18: /cgml_6/ptodata/2/pubppa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	67	100.0	12	US-10-218-743-5	Sequence 5, Appl
2	67	100.0	536	US-10-218-743-21	Sequence 21, Appl
3	67	100.0	555	US-10-218-743-15	Sequence 15, Appl
4	67	100.0	555	US-10-218-743-18	Sequence 18, Appl
5	62	92.5	490	US-10-218-743-41	Sequence 41, Appl
6	62	92.5	509	US-10-218-743-35	Sequence 35, Appl
7	62	92.5	509	US-10-218-743-38	Sequence 38, Appl
8	43	64.2	258	US-10-369-493-16621	Sequence 16621, A
9	41	61.2	251	US-10-282-122A-76697	Sequence 76697, A
10	40	59.7	236	US-09-738-626-6304	Sequence 6304, Ap
11	40	59.7	284	US-10-627-476-446	Sequence 446, App
12	40	59.7	284	US-10-627-476-450	Sequence 450, App
13	38	56.7	51	US-09-864-761-38990	Sequence 38990, A
14	38	56.7	359	US-09-838-955-3	Sequence 3, Appl
15	38	56.7	396	US-10-369-493-287	Sequence 287, App

16	37	55.2	53	12	US-10-424-599-180018	Sequence 180018, A
17	37	55.2	75	12	US-10-424-599-274388	Sequence 274388, A
18	37	55.2	86	12	US-10-424-599-243691	Sequence 243691, A
19	37	55.2	239	15	US-10-104-047-2759	Sequence 2759, Ap
20	37	55.2	326	15	US-10-094-749-2357	Sequence 2357, Ap
21	37	55.2	497	15	US-10-298-796-4	Sequence 4, Appl
22	37	55.2	575	9	US-09-938-405-2	Sequence 2, Appl
23	37	55.2	575	10	US-09-880-464-2	Sequence 2, Appl
24	37	55.2	575	14	US-10-150-440-3	Sequence 3, Appl
25	37	55.2	575	14	US-10-438-648-2	Sequence 2, Appl
26	37	55.2	575	15	US-10-094-986-196	Sequence 196, App
27	37	55.2	575	15	US-10-410-195-2	Sequence 2, Appl
28	37	55.2	575	15	US-10-373-801-29	Sequence 2, Appl
29	37	55.2	631	14	US-10-150-440-1	Sequence 1, Appl
30	37	55.2	686	14	US-10-417-719-25	Sequence 25, Appl
31	37	55.2	686	14	US-10-417-719-40	Sequence 40, Appl
32	37	55.2	686	14	US-10-417-719-42	Sequence 42, Appl
33	37	55.2	686	14	US-10-417-719-44	Sequence 44, Appl
34	37	55.2	686	14	US-10-417-719-46	Sequence 46, Appl
35	37	55.2	832	12	US-10-282-511-6	Sequence 6, Appl
36	37	55.2	847	12	US-10-282-511-8	Sequence 8, Appl
37	37	55.2	1013	12	US-10-425-114-62661	Sequence 62661, A
38	37	55.2	1198	12	US-10-210-172-26	Sequence 26, Appl
39	37	55.2	1286	12	US-09-964-956-38	Sequence 38, Appl
40	37	55.2	1286	13	US-10-017-216-7	Sequence 7, Appl
41	37	55.2	1441	15	US-10-412-897-3	Sequence 3, Appl
42	37	55.2	1597	12	US-09-964-956-41	Sequence 41, Appl
43	37	55.2	1597	13	US-10-017-216-6	Sequence 6, Appl
44	37	55.2	1641	12	US-09-964-956-40	Sequence 40, Appl
45	37	55.2	1641	13	US-10-017-216-5	Sequence 5, Appl

## ALIGNMENTS

```
RESULT 1
US-10-218-743-5
Sequence 5, Application US/10218743
Publicatoin No. US20030096779A1
GENERAL INFORMATION:
APPLICANT: McCall, Catherine A.
APPLICANT: Hunter, Shirley Wu
APPLICANT: Weber, Eric R.
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
FILE REFERENCE: AL-2-C3
CURRENT APPLICATION NUMBER: US/10/218, 743
CURRENT FILING DATE: 2002-08-13
PRIOR APPLICATION NUMBER: US/09/292,225
PRIOR FILING DATE: 1999-04-15
PRIOR APPLICATION NUMBER: 60/098,909
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/085,295
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/098,565
PRIOR FILING DATE: 1998-04-17
PRIOR APPLICATION NUMBER: 09/062,013
PRIOR FILING DATE: 1998-04-17
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 12
TYPE: PRT
ORGANISM: Dermatophagoides farinae
US-10-218-743-5
Query Match 100.0%; Score 67; DB 14; Length 12;
Best Local Similarity 100.0%; Pred. No. 1,7e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
RESULT 2
US-10-218-743-21
; Sequence 21, Application US/10218743
; Publication No. US20030096779A1
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/10/218,743
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US/09/292,225
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: 60/098,909
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/085,295
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/098,565
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: 09/062,013
; PRIOR FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-10-218-743-21

Query Match          100.0%; Score 67; DB 14; Length 536;
Best Local Similarity 100.0%; Pred. No. 0.00094;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DEKNSFECTILGP 12
Db      385 DEKNSFECTILGP 396

RESULT 3
US-10-218-743-15
; Sequence 15, Application US/10218743
; Publication No. US20030096779A1
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/10/218,743
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US/09/292,225
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: 60/098,909
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/085,295
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/098,565
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: 09/062,013
; PRIOR FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-10-218-743-15
```

```
Query Match          100.0%; Score 67; DB 14; Length 555;
Best Local Similarity 100.0%; Pred. No. 0.00098;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DEKNSFECTILGP 12
Db      404 DEKNSFECTILGP 415

RESULT 4
US-10-218-743-18
; Sequence 18, Application US/10218743
; Publication No. US20030096779A1
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/10/218,743
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US/09/292,225
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: 60/098,909
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/085,295
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/098,565
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: 09/062,013
; PRIOR FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-10-218-743-18

Query Match          100.0%; Score 67; DB 14; Length 555;
Best Local Similarity 100.0%; Pred. No. 0.00098;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DEKNSFECTILGP 12
Db      404 DEKNSFECTILGP 415

RESULT 5
US-10-218-743-41
; Sequence 41, Application US/10218743
; Publication No. US20030096779A1
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/10/218,743
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US/09/292,225
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: 60/098,909
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/085,295
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/098,565
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: 09/062,013
; PRIOR FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
```

SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 41  
LENGTH: 490  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-10-218-743-41

Query Match 92.5%; Score 62; DB 14; Length 490;  
Best Local Similarity 83.3%; Pred. No. 0.007;  
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 DEKNSPECILGP 12  
|||||:|:|  
Db 385 DEKNSYECILGP 396

RESULT 6  
US-10-218-743-35  
Sequence 35; Application US/10218743  
Publication No. US20030096779A1  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley Wu  
APPLICANT: Weber, Eric R.  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/10/218,743  
CURRENT FILING DATE: 2002-08-13  
PRIOR APPLICATION NUMBER: US/09/292,225  
PRIOR FILING DATE: 1999-04-15  
PRIOR APPLICATION NUMBER: 60/098,909  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/085,295  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/098,565  
PRIOR FILING DATE: 1998-04-17  
PRIOR APPLICATION NUMBER: 09/062,013  
PRIOR FILING DATE: 1998-04-17  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 35  
LENGTH: 509  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-10-218-743-35

Query Match 92.5%; Score 62; DB 14; Length 509;  
Best Local Similarity 83.3%; Pred. No. 0.0073;  
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 DEKNSPECILGP 12  
|||||:|:|  
Db 404 DEKNSYECILGP 415

RESULT 7  
US-10-218-743-38  
Sequence 38; Application US/10218743  
Publication No. US20030096779A1  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley Wu  
APPLICANT: Weber, Eric R.  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/10/218,743  
CURRENT FILING DATE: 2002-08-13  
PRIOR APPLICATION NUMBER: US/09/292,225  
PRIOR FILING DATE: 1999-04-15  
PRIOR APPLICATION NUMBER: 60/098,909  
PRIOR FILING DATE: 1998-09-02

PRIOR APPLICATION NUMBER: 60/085,295  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/098,565  
PRIOR FILING DATE: 1998-04-17  
PRIOR APPLICATION NUMBER: 09/062,013  
PRIOR FILING DATE: 1998-04-17  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 38  
LENGTH: 509  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-10-218-743-38

Query Match 92.5%; Score 62; DB 14; Length 509;  
Best Local Similarity 83.3%; Pred. No. 0.0073;  
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 DEKNSPECILGP 12  
|||||:|:|  
Db 404 DEKNSYECILGP 415

RESULT 8  
US-10-369-493-16621  
Sequence 16621; Application US/10369493  
Publication No. US2003023675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 16621  
LENGTH: 258  
TYPE: PRT  
ORGANISM: Bacillus thuringiensis  
US-10-369-493-16621

Query Match 64.2%; Score 43; DB 15; Length 258;  
Best Local Similarity 70.0%; Pred. No. 11;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 KNSPECILGP 12  
|||||:|:|  
Db 25 KNSYECILGP 34

RESULT 9  
US-10-282-122A-76697  
Sequence 76697; Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zykling, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Foreyth, R.  
APPLICANT: Xu, H.

```

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EUTRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 76697
; LENGTH: 291
; TYPE: PRT
; ORGANISM: Ureaplasma urealyticum
US-10-282-122A-76697

Query Match      61.2%; Score 41; DB 12; Length 291;
Best Local Similarity 72.7%; Pred. No. 28;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Cy 2 EKNSFECILGP 12
Db 53 KKNSTTILGP 63

RESULT 10
US-09-738-626-6304
; Sequence 6304, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6304
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum

```

```

US-09-738-626-6304

Query Match      59.7%; Score 40; DB 9; Length 236;
Best Local Similarity 63.6%; Pred. No. 35;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Cy 2 EKNSFECILGP 12
Db 25 KKNKFTALWGP 35

RESULT 11
US-10-627-476-446
; Sequence 446, Application US/10627476
; Publication No. US20040030116A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Mark
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schoder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauser, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
; TITLE OF INVENTION: INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE
; FILE REFERENCE: BGI-125CPN
; CURRENT APPLICATION NUMBER: US/10/627,476
; PRIOR FILING DATE: 2003-07-25
; PRIOR APPLICATION NUMBER: 09/602,787
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: USSN 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: DE 19931454.3
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931478.0
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931563.9
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19932122.1
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932124.8
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932125.6
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932128.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932180.9
; PRIOR FILING DATE: 1999-07-09
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 678
; SEQ ID NO 446
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-10-627-476-446

Query Match      59.7%; Score 40; DB 12; Length 284;
Best Local Similarity 63.6%; Pred. No. 42;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Cy 2 EKNSFECILGP 12
Db 73 KKNKFTALWGP 83

RESULT 12
US-10-627-476-450
; Sequence 450, Application US/10627476
; Publication No. US20040030116A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Mark
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schoder, Hartwig
; APPLICANT: Zelder, Oskar

```



APPLICANT: Haberhauser, Gregor  
TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS  
TITLE OF INVENTION: INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE  
TITLE OF INVENTION: TRANSPORT  
FILE REFERENCE: BGI-125PCPN  
CURRENT APPLICATION NUMBER: US/10/627,476  
CURRENT FILING DATE: 2003-07-25  
PRIOR APPLICATION NUMBER: 09/602,787  
PRIOR FILING DATE: 2000-06-23  
PRIOR APPLICATION NUMBER: USSN 60/141031  
PRIOR FILING DATE: 1999-06-25  
PRIOR APPLICATION NUMBER: DE 19931454.3  
PRIOR FILING DATE: 1999-07-08  
PRIOR APPLICATION NUMBER: DE 19931478.0  
PRIOR FILING DATE: 1999-07-08  
PRIOR APPLICATION NUMBER: DE 19931563.9  
PRIOR FILING DATE: 1999-07-08  
PRIOR APPLICATION NUMBER: DE 19932122.1  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: DE 19932124.8  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: DE 19932125.6  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: DE 19932128.0  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: DE 19932180.9  
PRIOR FILING DATE: 1999-07-09  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 678  
SEQ ID NO 450  
LENGTH: 284  
TYPE: PRT  
ORGANISM: Corynebacterium glutamicum  
US-10-627-476-450  
Query Match  
Best Local Similarity 59.7%; Score 40; DB 12; Length 284;  
Pred. No. 42;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 2 EXNSPECILGP 12  
Db 73 EXNKTALMGP 83  
RESULT 13  
US-09-864-761-38990  
Sequence 38990, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
FILE REFERENCE: Aecm1ca-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-06-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 38990  
LENGTH: 51  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC004824.1  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 12  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 12  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 11  
OTHER INFORMATION: EXPRESSED IN RETAL LIVER, SIGNAL = 13  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 10  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 9  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 12  
OTHER INFORMATION: SWISSPROT HIT: O88807, EVALUE 3.00e-07  
OTHER INFORMATION: EST\_HUMAN HIT: BE089551.1, EVALUE 2.90e+00  
US-09-864-761-38990  
Query Match  
Best Local Similarity 56.7%; Score 38; DB 9; Length 51;  
Pred. No. 16;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 1 DEXNSPELVGP 12  
Db 1 DNSTFELVGP 12  
RESULT 14  
US-09-838-955-3  
Sequence 3, Application US/09838955  
Patent No. US20020056152A1  
GENERAL INFORMATION:  
APPLICANT: Kelly, James D  
APPLICANT: Melotto, Maeji  
TITLE OF INVENTION: DNA Encoding For A Disease Resistance Gene From Common  
TITLE OF INVENTION: Bean and Methods of Use  
FILE REFERENCE: 6550-000044  
CURRENT APPLICATION NUMBER: US/09/838,955  
CURRENT FILING DATE: 2001-04-20  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 3  
LENGTH: 369  
TYPE: PRT  
ORGANISM: Phaseolus vulgaris  
US-09-838-955-3  
Query Match  
Best Local Similarity 56.7%; Score 38; DB 9; Length 369;  
Pred. No. 1.3e+02;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 EKNSPECTIGP 12  
Db 162 KRSTFHCTIGP 172

## RESULT 15

US-10-369-493-287  
; Sequence 287, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianning  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 287  
; LENGTH: 396  
; TYPE: PRT  
; ORGANISM: Xenorhabdus nematophilus  
US-10-369-493-287

Query Match 56.7%; Score 38; DB 15; Length 396;  
Best Local Similarity 54.5%; Pred. No. 1.4e+02;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 DEKNSPECTIG 11  
Db 264 DDKETKECVLIG 274

Search completed: March 22, 2004, 07:45:48  
Job time : 2.4547 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:42:54 ; Search time 0.967204 Seconds  
(without alignments)  
640,518 Million cell updates/sec

Title: US-09-662-293-5  
Perfect score: 67  
Sequence: 1 DEKNSFECILGP 12

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/iaa/5A.COMB.pep:\*  
2: /cgn2\_6/prodata/2/iaa/5B.COMB.pep:\*  
3: /cgn2\_6/prodata/2/iaa/5A.COMB.pep:\*  
4: /cgn2\_6/prodata/2/iaa/5B.COMB.pep:\*  
5: /cgn2\_6/prodata/2/iaa/5A.COMB.pep:\*  
6: /cgn2\_6/prodata/2/iaa/5B.COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	67	100.0	12	US-09-292-225-5	Sequence 5, Appl
2	67	100.0	536	US-09-292-225-21	Sequence 21, Appl
3	67	100.0	555	US-09-292-225-15	Sequence 15, Appl
4	67	100.0	555	US-09-292-225-18	Sequence 18, Appl
5	62	92.5	490	US-09-292-225-41	Sequence 41, Appl
6	62	92.5	509	US-09-292-225-35	Sequence 35, Appl
7	62	92.5	509	US-09-292-225-38	Sequence 38, Appl
8	39	58.2	118	US-09-328-352-38	Sequence 38, Appl
9	39	58.2	2703	US-08-185-432-19	Sequence 5794, Ap
10	39	58.2	2703	US-08-899-232-4	Sequence 19, Appl
11	38	56.7	369	US-09-838-955A-3	Sequence 4, Appl
12	38	56.7	609	US-08-927-219-129	Sequence 129, Appl
13	37.5	56.0	166	US-08-818-723-4	Sequence 4, Appl
14	37.5	56.0	166	US-09-237-507-4	Sequence 4, Appl
15	37.5	56.0	166	US-08-261-206A-4	Sequence 4, Appl
16	37	55.2	114	US-08-733-564-1	Sequence 9, Appl
17	37	55.2	115	US-08-312-870-9	Sequence 9, Appl
18	37	55.2	145	US-09-252-991A-22680	Sequence 22680, A
19	37	55.2	275	US-08-512-870-7	Sequence 7, Appl
20	37	55.2	456	US-08-307-444A-3	Sequence 3, Appl
21	37	55.2	456	US-08-307-444A-4	Sequence 4, Appl
22	37	55.2	456	US-08-307-444A-3	Sequence 4, Appl
23	37	55.2	456	US-08-307-444A-4	Sequence 4, Appl
24	37	55.2	456	US-08-307-444A-1	Sequence 1, Appl
25	37	55.2	475	US-08-307-444A-2	Sequence 2, Appl
26	37	55.2	475	US-08-307-444A-1	Sequence 1, Appl
27	37	55.2	475	US-08-307-444A-2	Sequence 2, Appl

28	37	55.2	476	US-08-014-723-1	Sequence 1, Appl
29	37	55.2	476	US-08-014-723-2	Sequence 2, Appl
30	37	55.2	476	US-08-014-723-18	Sequence 18, Appl
31	37	55.2	476	US-08-110-011A-1	Sequence 1, Appl
32	37	55.2	476	US-08-110-011A-2	Sequence 2, Appl
33	37	55.2	476	US-08-110-011A-18	Sequence 18, Appl
34	37	55.2	494	US-08-014-723-14	Sequence 14, Appl
35	37	55.2	494	US-08-014-723-16	Sequence 16, Appl
36	37	55.2	494	US-08-110-011A-14	Sequence 14, Appl
37	37	55.2	494	US-08-110-011A-16	Sequence 16, Appl
38	37	55.2	497	US-08-312-870-3	Sequence 3, Appl
39	37	55.2	497	US-09-331-793-4	Sequence 4, Appl
40	37	55.2	498	US-08-733-564-2	Sequence 2, Appl
41	37	55.2	572	US-08-733-564-2	Sequence 2, Appl
42	37	55.2	575	US-08-261-206A-59	Sequence 59, Appl
43	37	55.2	575	US-08-312-870-1	Sequence 1, Appl
44	37	55.2	575	US-08-170-290A-54	Sequence 54, Appl
45	37	55.2	575	US-09-880-484D-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-09-292-225-5  
Sequence 5, Application US/09292225  
Patent No. 6455686  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Weber, Eric R.  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/09/292,225  
CURRENT FILING DATE: 1999-04-15  
EARLIER APPLICATION NUMBER: 60/098,909  
EARLIER FILING DATE: 1998-09-02  
EARLIER APPLICATION NUMBER: 60/085,295  
EARLIER FILING DATE: 1998-05-13  
EARLIER APPLICATION NUMBER: 60/098,565  
EARLIER FILING DATE: 1998-04-17  
EARLIER APPLICATION NUMBER: 09/062,013  
EARLIER FILING DATE: 1998-04-17  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 5  
LENGTH: 12  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-09-292-225-5

Query Match 100.0%; Score 67; DB 4; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.2e-05;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 1 DEKNSFECILGP 12  
1 DEKNSFECILGP 12

RESULT 2  
US-09-292-225-21  
Sequence 21, Application US/09292225  
Patent No. 6455686  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley Wu  
APPLICANT: Weber, Eric R.  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/09/292,225

;; CURRENT FILING DATE: 1999-04-15  
;; EARLIER APPLICATION NUMBER: 60/098,909  
;; EARLIER FILING DATE: 1998-09-02  
;; EARLIER APPLICATION NUMBER: 60/085,295  
;; EARLIER FILING DATE: 1998-05-13  
;; EARLIER APPLICATION NUMBER: 60/098,565  
;; EARLIER FILING DATE: 1998-04-17  
;; EARLIER APPLICATION NUMBER: 09/062,013  
;; EARLIER FILING DATE: 1998-04-17  
;; NUMBER OF SEQ ID NOS: 49  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 21  
;; LENGTH: 536  
;; TYPE: PRT  
;; ORGANISM: Dermatophagoides farinae  
US-09-292-225-21

Query Match 100.0%; Score 67; DB 4; Length 536;  
Best Local Similarity 100.0%; Pred. NO. 0.00058;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DEKNSFECTILGP 12  
Db 385 DEKNSFECTILGP 396

RESULT 3  
US-09-292-225-15  
;; Sequence 15, Application US/09292225  
;; Patent No. 6455686  
;; GENERAL INFORMATION:  
;; APPLICANT: McCall, Catherine A.  
;; APPLICANT: Hunter, Shirley Wu  
;; APPLICANT: Weber, Eric R.  
;; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
;; FILE REFERENCE: AL-2-C3  
;; CURRENT APPLICATION NUMBER: US/09/292,225  
;; CURRENT FILING DATE: 1999-04-15  
;; EARLIER APPLICATION NUMBER: 60/098,909  
;; EARLIER FILING DATE: 1998-09-02  
;; EARLIER APPLICATION NUMBER: 60/085,295  
;; EARLIER FILING DATE: 1998-05-13  
;; EARLIER APPLICATION NUMBER: 60/098,565  
;; EARLIER FILING DATE: 1998-04-17  
;; EARLIER APPLICATION NUMBER: 09/062,013  
;; EARLIER FILING DATE: 1998-04-17  
;; NUMBER OF SEQ ID NOS: 49  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 15  
;; LENGTH: 555  
;; TYPE: PRT  
;; ORGANISM: Dermatophagoides farinae  
US-09-292-225-15

Query Match 100.0%; Score 67; DB 4; Length 555;  
Best Local Similarity 100.0%; Pred. NO. 0.0006;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DEKNSFECTILGP 12  
Db 404 DEKNSFECTILGP 415

RESULT 4  
US-09-292-225-18  
;; Sequence 18, Application US/09292225  
;; Patent No. 6455686  
;; GENERAL INFORMATION:  
;; APPLICANT: McCall, Catherine A.  
;; APPLICANT: Hunter, Shirley Wu  
;; APPLICANT: Weber, Eric R.  
;; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS

;; TITLE OF INVENTION: AND USES THEREOF  
;; FILE REFERENCE: AL-2-C3  
;; CURRENT APPLICATION NUMBER: US/09/292,225  
;; CURRENT FILING DATE: 1999-04-15  
;; EARLIER APPLICATION NUMBER: 60/098,909  
;; EARLIER FILING DATE: 1998-09-02  
;; EARLIER APPLICATION NUMBER: 60/085,295  
;; EARLIER FILING DATE: 1998-05-13  
;; EARLIER APPLICATION NUMBER: 60/098,565  
;; EARLIER FILING DATE: 1998-04-17  
;; EARLIER APPLICATION NUMBER: 09/062,013  
;; EARLIER FILING DATE: 1998-04-17  
;; NUMBER OF SEQ ID NOS: 49  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 18  
;; LENGTH: 555  
;; TYPE: PRT  
;; ORGANISM: Dermatophagoides farinae  
US-09-292-225-18

Query Match 100.0%; Score 67; DB 4; Length 555;  
Best Local Similarity 100.0%; Pred. NO. 0.0006;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DEKNSFECTILGP 12  
Db 404 DEKNSFECTILGP 415

RESULT 5  
US-09-292-225-41  
;; Sequence 41, Application US/09292225  
;; Patent No. 6455686  
;; GENERAL INFORMATION:  
;; APPLICANT: McCall, Catherine A.  
;; APPLICANT: Hunter, Shirley Wu  
;; APPLICANT: Weber, Eric R.  
;; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
;; FILE REFERENCE: AL-2-C3  
;; CURRENT APPLICATION NUMBER: US/09/292,225  
;; CURRENT FILING DATE: 1999-04-15  
;; EARLIER APPLICATION NUMBER: 60/098,909  
;; EARLIER FILING DATE: 1998-09-02  
;; EARLIER APPLICATION NUMBER: 60/085,295  
;; EARLIER FILING DATE: 1998-05-13  
;; EARLIER APPLICATION NUMBER: 60/098,565  
;; EARLIER FILING DATE: 1998-04-17  
;; EARLIER APPLICATION NUMBER: 09/062,013  
;; EARLIER FILING DATE: 1998-04-17  
;; NUMBER OF SEQ ID NOS: 49  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 41  
;; LENGTH: 490  
;; TYPE: PRT  
;; ORGANISM: Dermatophagoides farinae  
US-09-292-225-41

Query Match 92.5%; Score 62; DB 4; Length 490;  
Best Local Similarity 83.3%; Pred. NO. 0.004;  
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 DEKNSFECTILGP 12  
Db 385 DEKNSFECTILGP 396

RESULT 6  
US-09-292-225-35  
;; Sequence 35, Application US/09292225  
;; Patent No. 6455686  
;; GENERAL INFORMATION:  
;; APPLICANT: McCall, Catherine A.

APPLICANT: Hunter, Shirley Wu  
APPLICANT: Weber, Eric R.  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
TITLE OF INVENTION: AND USES THEREOF  
FILE REFERENCE: AL-2-C3  
CURRENT FILING DATE: 1999-04-15  
CURRENT APPLICATION NUMBER: US/09/292,225  
EARLIER FILING DATE: 1998-09-02  
EARLIER APPLICATION NUMBER: 60/098,909  
EARLIER FILING DATE: 1998-05-13  
EARLIER APPLICATION NUMBER: 60/098,565  
EARLIER FILING DATE: 1998-04-17  
EARLIER APPLICATION NUMBER: 09/062,013  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 35  
LENGTH: 509  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-09-292-225-35

Query Match 92.5%; Score 62; DB 4; Length 509;  
Best Local Similarity 83.3%; Pred. No. 0.0042;  
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 DEKNSPECILGP 12  
|||||:|||||  
Db 404 DEKNSYECLIGP 415

US-09-292-225-38  
Sequence 38, Application US/09292225  
Patent No. 6455866  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley Wu  
APPLICANT: Weber, Eric R.  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
TITLE OF INVENTION: AND USES THEREOF  
FILE REFERENCE: AL-2-C3  
CURRENT FILING DATE: 1999-04-15  
EARLIER FILING DATE: 1998-09-02  
EARLIER APPLICATION NUMBER: 60/098,909  
EARLIER FILING DATE: 1998-09-02  
EARLIER APPLICATION NUMBER: 60/085,295  
EARLIER FILING DATE: 1998-05-13  
EARLIER APPLICATION NUMBER: 60/098,565  
EARLIER FILING DATE: 1998-04-17  
EARLIER APPLICATION NUMBER: 09/062,013  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 38  
LENGTH: 509  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-09-292-225-38

Query Match 92.5%; Score 62; DB 4; Length 509;  
Best Local Similarity 83.3%; Pred. No. 0.0042;  
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 DEKNSPECILGP 12  
|||||:|||||  
Db 404 DEKNSYECLIGP 415

RESULT 8  
US-09-328-352-5794  
Sequence 5794, Application US/09328352

Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Gary L. Bregon et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC99-03PA  
CURRENT FILING DATE: 1999-06-04  
CURRENT APPLICATION NUMBER: US/09/328,352  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 5794  
LENGTH: 118  
TYPE: PRT  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-5794

Query Match 58.2%; Score 39; DB 4; Length 118;  
Best Local Similarity 54.5%; Pred. No. 11;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 DEKNSPECILGP 11  
|||||:|||||  
Db 48 DRGVYECVIG 58

RESULT 9  
US-08-185-432-19  
Sequence 19, Application US/08185432  
Patent No. 5750652  
GENERAL INFORMATION:  
APPLICANT: Attavanis-Tsakonas, Spyridon  
APPLICANT: Buseau, Isabelle  
APPLICANT: Diederich, Robert J.  
APPLICANT: Xu, Tian  
APPLICANT: Matsuno, Kenji  
TITLE OF INVENTION: DELTA PROTEINS, NUCLEIC ACIDS, AND  
NUMBER OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT FILING DATE: 21-JAN-1994  
CURRENT APPLICATION NUMBER: US/08/185,432  
FILING DATE: 21-JAN-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Mierlock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-8090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2703 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-185-432-19

Query Match 58.2%; Score 39; DB 4; Length 2703;  
Best Local Similarity 50.0%; Pred. No. 2.6e+02;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DEKNSPECTILGP 12  
DB 618 DVNSFKCLDP 629

RESULT 10  
US-08-899-232-4  
Sequence 4, Application US/08899232

PATENT INFORMATION:  
PATENT NO. 6436650  
APPLICANT: Artavanis-Teakonas, Spyridon  
TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON  
FILE REFERENCE: 7326-046  
CURRENT APPLICATION NUMBER: US/08/899,232  
CURRENT FILING DATE: 1997-07-23  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 4  
LENGTH: 2703  
TYPE: PRT  
ORGANISM: Drosophila sp.  
US-08-899-232-4

Query Match 58.2%; Score 39; DB 4; Length 2703;  
Best Local Similarity 50.0%; Pred. No. 2.6e+02;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DEKNSPECTILGP 12  
DB 618 DVNSFKCLDP 629

RESULT 11  
US-09-838-955A-3  
Sequence 3, Application US/09838955A  
Patent No. 6646183  
GENERAL INFORMATION:  
APPLICANT: Michigan State University  
APPLICANT: Kelly, James D  
TITLE OF INVENTION: DNA ENCODING FOR A DISEASE RESISTANCE GENE FROM COMMON BEAN AND M  
TITLE OF INVENTION: OF USE  
FILE REFERENCE: 6550-000044  
CURRENT APPLICATION NUMBER: US/09/838,955A  
CURRENT FILING DATE: 2003-02-24  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 3  
LENGTH: 369  
TYPE: PRT  
ORGANISM: Phaseolus vulgaris  
US-09-838-955A-3

Query Match 56.7%; Score 38; DB 4; Length 369;  
Best Local Similarity 54.5%; Pred. No. 52;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 EKNSPECTILGP 12  
DB 162 KRSTFHCLDP 172

RESULT 12  
US-08-927-219-129  
Sequence 129, Application US/08927219  
Patent No. 6187533  
GENERAL INFORMATION:  
APPLICANT: Bell, Graeme I.  
APPLICANT: Yamagata, Kazuya  
APPLICANT: Oda, Naohisa  
APPLICANT: Katsaki, Pamela J.  
APPLICANT: Fututa, Hiroto

APPLICANT: Horikawa, Yukio  
APPLICANT: Menzel, Stephen  
TITLE OF INVENTION: MUTATIONS IN THE DIABETES SUSCEPTIBILITY  
TITLE OF INVENTION: GENES HEPATOCYTE NUCLEAR FACTOR (HNF) 1 ALPHA, HNF-1BETA  
TITLE OF INVENTION: AND HNF-4ALPHA  
NUMBER OF SEQUENCES: 147  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/927,219  
FILING DATE: Concurrently Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/029,679  
FILING DATE: 30-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/028,056  
FILING DATE: 02-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/025,719  
FILING DATE: 10-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Wilson, Mark B. 37,259  
REGISTRATION NUMBER: 37,259  
REFERENCE/DOCKET NUMBER: ARCD:272  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 512/418-3000  
TELEFAX: 512/474-7577  
INFORMATION FOR SEQ ID NO: 129:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 609 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-927-219-129

Query Match 56.7%; Score 38; DB 3; Length 609;  
Best Local Similarity 77.8%; Pred. No. 86;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 NSPECTILGP 12  
DB 381 NAFEMILGP 389

RESULT 13  
US-08-918-723-4  
Sequence 4, Application US/08918723  
Patent No. 5863779  
GENERAL INFORMATION:  
APPLICANT: Lal, Preeti  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: UBC7-LIKE UBIQUITIN-CONJUGATING ENZYME  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette

```

1  COMPUTER: IBM Compatible
2  OPERATING SYSTEM: DOS
3  SOFTWARE: FASTSEQ for Windows Version 2.0
4  CURRENT APPLICATION DATA:
5  APPLICATION NUMBER: US/08/918,723
6  FILING DATE: Herewith
7  CLASSIFICATION: 435
8  PRIOR APPLICATION DATA:
9  APPLICATION NUMBER:
10 FILING DATE:
11 ATTORNEY/AGENT INFORMATION:
12 NAME: Billings, Lucy J.
13 REGISTRATION NUMBER: 36,749
14 REFERENCE/DOCKET NUMBER: PF-0368 US
15 TELECOMMUNICATION INFORMATION:
16 TELEPHONE: 415-855-0555
17 TELEFAX: 415-845-4166
18 TELEX:
19 INFORMATION FOR SEQ ID NO: 4:
20 SEQUENCE CHARACTERISTICS:
21 LENGTH: 166 amino acids
22 TYPE: amino acid
23 STRANDEDNESS: single
24 TOPOLOGY: linear
25 IMMEDIATE SOURCE:
26 LIBRARY: GenBank
27 CLONE: 992704
28 US-08-918-723-4
29
30 Query Match 56.0% Score 37.5; DB 2; Length 166;
31 Best Local Similarity 60.0% Pred. No. 28;
32 Matches 9; Conservative 1; Mismatches 2; Indels 3; Gaps 1;
33
34 QY 1 DEKNSFE--CILGP 12
35 ||||| |||
36 ||||| |||
37 ||||| |||
38 ||||| |||
39 Db 30 DEKNFEKSVTLIGP 44
40
41 RESULT 14
42 US-09-237-507-4
43 Sequence 4, Application US/09237507
44 Patent No. 5952181
45
46 GENERAL INFORMATION:
47 APPLICANT: Lal, Preeti
48 APPLICANT: Corley, Neil C.
49 TITLE OF INVENTION: UBC7-LIKE UBIQUITIN-CONJUGATING ENZYME
50 NUMBER OF SEQUENCES: 4
51 CORRESPONDENCE ADDRESS:
52 ADDRESSEE: Incyte Pharmaceuticals, Inc.
53 STREET: 3174 Porter Drive
54 CITY: Palo Alto
55 STATE: CA
56 COUNTRY: USA
57 ZIP: 94304
58
59 COMPUTER READABLE FORM:
60 MEDIUM TYPE: Diskette
61 COMPUTER: IBM Compatible
62 OPERATING SYSTEM: DOS
63 SOFTWARE: FASTSEQ for Windows Version 2.0
64 CURRENT APPLICATION DATA:
65 APPLICATION NUMBER: US/09/237,507
66 FILING DATE:
67 CLASSIFICATION:
68 PRIOR APPLICATION DATA:
69 APPLICATION NUMBER: 08/918,723
70 FILING DATE:
71 ATTORNEY/AGENT INFORMATION:
72 NAME: Billings, Lucy J.
73 REGISTRATION NUMBER: 36,749
74 REFERENCE/DOCKET NUMBER: PF-0368 US
75 TELECOMMUNICATION INFORMATION:
76 TELEPHONE: 415-855-0555
77 TELEFAX: 415-845-4166
78

```

```

TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 166 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Genbank
; CLONE: 992704
US-09-237-507-4

Query Match      56.0%: Score 37.5; DB 2; Length 166;
Best Local Similarity    60.0%; Pred. No. 28;
Matches   9; Conservative   1; Mismatches   2; Indels   3; Gaps   1.

QY          1 DEKNSFPR--CITGP 12
            ||||| | :|||
Db          30 DEKNIFESVITIGP 44

RESULT 15
US-08-261-206A-4
; Sequence 4, Application US/08261206A
; Patent No. 5574007
; GENERAL INFORMATION:
; APPLICANT: Zushi, Mitichitaka
; APPLICANT: Gomi, Komakazu
; APPLICANT: Yamamoto, Shuji
; APPLICANT: Suzuki, Koji
; APPLICANT: Matsuda, Akio
; TITLE OF INVENTION: A Polypeptide Capable of Interacting
; TITLE OF INVENTION: with Thrombin
; NUMBER OF SEQUENCES: 80
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 301 N. Washington St.
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22046-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,206A
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/740,492
; FILING DATE: 03-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Svensson, Leonard R.
; REGISTRATION NUMBER: 30330
; REFERENCE/DOCKET NUMBER: 216-275P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-241-1300
; TELEFAX: 703-241-2848
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 36 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: C-terminal
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..36
; OTHER INFORMATION: /label= peptide

```

OTHER INFORMATION: /note= "Peptide II, preferred peptide to be  
; OTHER INFORMATION: attached to C-terminus of thrombin binding  
; OTHER INFORMATION: polypeptide, Peptide I."  
US-08-261-206A-4

Query Match 55.2%; Score 37; DB 1; Length 36;  
Best Local Similarity 75.0%; Pred. No. 7.3;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 SPECILGP 12  
: ||| |  
Db 17 TPECICGP 24

Search completed: March 22, 2004, 07:03:55  
Job time : 1.9672 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:40:28 ; Search time 1.17065 Seconds

(without alignments)  
1479.047 Million cell updates/sec

Title: US-09-662-293-6

Perfect score: 96

Sequence: 1 DAFPHGYLLTAASPSK 18

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 78: \*  
2: PIR1: \*  
3: PIR2: \*  
4: PIR3: \*  
5: PIR4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	65.6	1635	2 T14075	chitinase (EC 3.2.
2	45.5	47.4	258	2 G65054	hypothetical prote
3	44	45.8	395	2 T39956	probable nadh-depe
4	44	45.8	605	2 D69149	hypothetical prote
5	44	45.8	636	2 F72736	hypothetical prote
6	44	45.8	1611	2 T38236	hypothetical prote
7	43	44.8	88	2 A69733	PBSX prophage ORF
8	43	44.8	265	2 E72279	hypothetical prote
9	43	44.8	425	2 H90415	hypothetical prote
10	43	44.8	655	2 G75582	cell division prot
11	42.5	44.3	452	2 T20556	hypothetical prote
12	42	43.8	104	1 GMB0	gastrointestinal prote
13	42	43.8	255	2 T10820	photosynthetic rea
14	42	43.8	420	2 B84111	sugar transport sy
15	42	43.8	449	2 T22777	hypothetical prote
16	42	43.8	500	2 A43656	sigk-creating site
17	41	42.7	57	2 T49863	hypothetical prote
18	41	42.7	276	2 A28170	reaction center pr
19	41	42.7	284	2 B84435	probable phloem-sp
20	41	42.7	378	2 H69280	NADH-dependent fla
21	41	42.7	395	2 C97244	NADH:flavin oxidor
22	41	42.7	418	2 S74875	chitinase (EC 3.2.
23	41	42.7	483	2 A53918	hypothetical prote
24	41	42.7	494	2 T32685	hypothetical prote
25	41	42.7	1184	2 H86190	MEGF2 protein - hu
26	41	42.7	1364	2 T00250	alpha-2-macroglobu
27	41	42.7	1507	2 T18544	hypothetical prote
28	40.5	42.2	231	2 AE2433	hypothetical prote
29	40	41.7	173	2 T00650	hypothetical prote

30	40	41.7	209	2 C87617	glutathione S-tran
31	40	41.7	227	2 AP2086	hypothetical prote
32	40	41.7	231	2 B71298	hypothetical prote
33	40	41.7	253	2 T08175	chlorophyll a/b-bi
34	40	41.7	262	2 T05654	hypothetical prote
35	40	41.7	451	2 AF3053	conserved hypothet
36	40	41.7	451	2 G98232	hypothetical prote
37	40	41.7	459	2 F71131	probable methyltra
38	40	41.7	471	2 T49019	probable RNA bindi
39	40	41.7	525	2 T44445	chitinase (EC 3.2.
40	40	41.7	545	2 AD2062	hypothetical prote
41	40	41.7	716	1 P2IV34	RNA-directed RNA p
42	40	41.7	716	1 P2IVMS	RNA-directed RNA p
43	40	41.7	716	1 P2IV68	RNA-directed RNA p
44	40	41.7	716	1 P2IV61	RNA-directed RNA p
45	40	41.7	725	2 T08989	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

T14075

chitinase (EC 3.2.1.14) - yellow fever mosquito

C/Species: Aedes aegypti (yellow fever mosquito)

C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C/Accession: T14075

Ride la Vega, H.; Specht, C.A.; Liu, Y.; Robbins, P.W.

Insect Mol. Biol. 7, 233-239, 1997

A/Title: Chitinases are a multi-gene family in Aedes, Anopheles, and Drosophila.

A/Reference number: 217872

A/Accession: T14075

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1635 <DBL>

A/Cross-references: EMBL:AF026492; NID:g2564720; PID:g2564721; PID:AA81850.1

C/Genetics:

A/Gene: CHT2

A/Introns: 462/3; 524/3; 618/1; 951/3; 1151/2

C/Keywords: glycosidase; hydrolase; polysaccharide degradation

##### Query Match

Best Local Similarity 65.6%; Score 63; DB 2; Length 1635;

Matches 12; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 AEFPHGYLLTAASPSK 18

DB 393 AFKPRGWLTAASPSK 409

##### RESULT 2

G65054

hypothetical protein b2739 - Escherichia coli (strain K-12)

C/Species: Escherichia coli

C/Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002

C/Accession: G65054

R. Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C

.A.; Rose, D.J.; Mau, B.; Sho, Y.

Science 277, 1453-1462, 1997

A/Title: The complete genome sequence of Escherichia coli K-12.

A/Reference number: A64720; MUID:97426617; PMID:9278503

A/Accession: G65054

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Cross-references: GB:AF000357; GB:U00096; NID:g2367155; PID:AACT5781.1; PID:g1789095

A/Experimental source: Strain K-12, substrain MG1655

C/Superfamily: conserved hypothetical protein H11013

Query Match 47.4%; Score 45.5; DB 2; Length 258;

Best Local Similarity 55.6%; Pred. No. 6.7;

Matches 10; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 1 DAFEPHC-YLITAAVSPG 17  
 Db 131 DRFAPHGKRLIVLSPG 148

## RESULT 3

T39956  
 probable nadh-dependent flavin oxidoreductase - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 18-Feb-2000  
 C:Accession: T39956  
 R:Xiang, Z.; Aves, S.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
 submitted to the EMBL Data Library, January 1999  
 A:Reference number: Z21893  
 A:Accession: T39956  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-395 <X1A>  
 A:Cross-references: EMBL:AL035065; PIDN:CAA22626.1; GSPDB:GN00067; SPDB:SPBC23G7.10c  
 A:Experimental source: strain 972h-, cosmid c23G7  
 C:Genetics:  
 A:Gene: SPDB:SPBC23G7.10c  
 A:Map position: 2  
 C:Superfamily: NADPH dehydrogenase chain OYE2

Query Match 45.8%; Score 44; DB 2; Length 395;  
 Best Local Similarity 63.6%; Pred. No. 18;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 6 HGYLITAAVSP 16  
 |||||  
 Db 211 HGYLITVSP 221

## RESULT 4

D69149  
 Hypothetical protein MTH381 - Methanobacterium thermoautotrophicum (strain Delta H)  
 C:Species: Methanobacterium thermoautotrophicum  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
 C:Accession: D69149  
 R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;  
 Olt, D.; Spatafora, R.; Viscate, R.; Wang, Y.; Mierzowski, J.; Gibson, R.; Jivani, N.;  
 Ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Neelling, J.; Reeve, J.N.  
 J. Bacteriol. 179, 7135-7155, 1997  
 A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct  
 A:Reference number: A69000; MUID:98037514; PMID:9371463  
 A:Accession: D69149  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-605 <MTH>  
 A:Cross-references: GB:A6000623; GB:A600066; NID:G2621432; PIDN:AAE84887.1; PID:G262144  
 A:Experimental source: strain Delta H  
 C:Genetics:  
 A:Gene: MTH381  
 A:Start codon: GTG

Query Match 45.8%; Score 44; DB 2; Length 605;  
 Best Local Similarity 80.0%; Pred. No. 29;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 7 GYLITAAVSP 16  
 |||||  
 Db 377 GYLITAAVAP 386

## RESULT 5

F72736  
 Hypothetical protein APE0428 - Aeropyrum pernix (strain K1)  
 C:Species: Aeropyrum pernix  
 C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
 C:Accession: F72736  
 R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Hakiwa, Y.; Jin-no, K.; Takai  
 awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, U.; N

DNA Res. 6, 83-101, 1999  
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr  
 A:Reference number: A72450; MUID:99310339; PMID:10382966  
 A:Accession: F72736  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-636 <KAW>  
 A:Cross-references: DBJ:AP000059; NID:G5103911; PIDN:BAF79386.1; PID:dl043172; PID:G510  
 A:Experimental source: strain K1  
 C:Genetics:  
 A:Gene: APE0428

Query Match 45.8%; Score 44; DB 2; Length 636;  
 Best Local Similarity 50.0%; Pred. No. 30;  
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 6 HGYLITAAVSPG 17  
 |||||  
 Db 163 YGVITGALTGP 174

## RESULT 6

T38236  
 Hypothetical protein SPAC23A1.17 - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
 C:Accession: T38236  
 R:Murphy, L.; Harris, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.  
 submitted to the EMBL Data Library, February 1998  
 A:Reference number: Z21780  
 A:Accession: T38236  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1611 <MUR>  
 A:Cross-references: EMBL:AL021813; PIDN:CAA16991.1; GSPDB:GN00066; SPDB:SPAC23A1.17  
 A:Experimental source: strain 972h-, cosmid c23A1  
 C:Genetics:  
 A:Gene: SPDB:SPAC23A1.17  
 A:Map position: 1

Query Match 45.8%; Score 44; DB 2; Length 1611;  
 Best Local Similarity 50.0%; Pred. No. 80;  
 Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 DAFEPHGKRLIVLSPG 16  
 |||||  
 Db 853 DFEYPHSTYLESPAPFP 868

## RESULT 7

A69733  
 PBX prophage ORF xkdr - Bacillus subtilis  
 C:Species: Bacillus subtilis  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999  
 C:Accession: A69733  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berr  
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho  
 A.; Ehrlich, S.D.; Emmerson, P.T.; Eutlian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall  
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulo, M.F.  
 Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
 A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue  
 Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portelle  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon  
 A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Ser  
 alexuch, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
 A.; Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A:Reference number: A69580; MUID:98044033; PMID:9384377  
 A:Accession: A69733  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
A:Residues: 1-88 <KUN>  
A:Cross-references: GB:Z99110, GB:AI009126, NID:92633472, PIDN:CB813128.1, PID:e1183291  
C:Genetics:  
A:Experimental source: strain 168  
A:Gene: xkdr

Query Match	44.8%;	Score 43;	DB 2;	Length 88;
Best Local Similarity	50.0%;	Pred. No. 5.7;		
Matches	9;	Conservative	4;	Mismatches 5;
				Indels 0;
				Gaps 0;

QY	1	DAFEPHGYLLTAAVSPGK	18
		: : :	:
Ds	63	DALEPGDRLMTAALTGCG	80

RESULT 8  
B72279  
hypochemical protein TM123 - Thermotoga maritima (strain MS8)  
CISpecies: Thermotoga maritima  
CDate: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
CAccession: B72279  
R:Neilson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
R:Neilson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garte, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.  
C.M.

Nature 399, 323-329, 1999  
 A>Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
 A.Reference number: A72200; PMID:99287316; PMID:10360571

A/Accession: B72279  
A/Status: preliminary  
A/Molecule type: DNA

A:Residues: 1-265 <ARN>  
A:Cross-references: GB:AE001779; GB:AE00512; NID:g4981777; PID:g4981777  
C:Experimental source: strain MSB8

C/Genetics:  
A/Gene: TM1233  
C/Superfamily: maltose transport protein malG

Query Match	44.8%	Score 43;	DB 2;	Length 265;
Best Local Similarity	64.3%	Pred. No. 18;		
Matches	9;	Conservative	0;	Mismatches 5;
				Indels 0;
				Gaps 0;

QY 5 PHGYLLTAAVSPGK 18  
|||  
Db 85 PAGYALTRYVFPCK 98

```

RESULT 9
H90415
hypothetical protein glna-2 [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #ext_change 24-May-2001

```

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Mavez, M.T.; Chan Jong, I.; Jeffries, A.C.; Kozera, C.U.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R.; R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.  
submitted to Genbank, April 2001  
A:Description: *Sulfolobus solfataricus* complete genome.

A/Accession: H90415  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1425 <KIR>  
A/Cross-references: GB:AE006641; NID:g13815744; PIDD:AAK4283.1; GSPDB:GN001555  
C/Genetics:  
C/Gene: g1na-2

```

Query Match      44.8%  Score 43:  DB 2;  length 425;
Best Local Similarity 53.3%  Pred. No. 29;
Matches      8;  Conservative      3;  Indels      0;  Gaps      0;
Oy          2  AFEPHGYLLTAAVSP 16
      :||| ||| :|:|

```

Db 130 SEPTFYLLNSALNP 144

RESULT 10  
G75582

Cell division protein FtsH - Deinococcus radiodurans (Strain R1)  
C/Spectres: Deinococcus radiodurans  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
C/Accession: G75582  
R/Miller, O.: Eisen, J.A.: Heidelberg, J.F.: Hickey, E.K.: Peterson, J.D.: Dodson, R.: Shen, M.: Vamathevan, J.: Lam, P.: McDonald, L.: Utterback, T.: Zalewski, C.: Smith, H.O.: Venter, J.C.: Fraser, C.M.  
Science 286, 1571-1577, 1999

**At**Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R.  
**A**Reference number: A75250; MUID:20036896; PMID:10567266  
**A**Accession: G75582

```

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-655 <MM1>

```

A:Cross-references: GB:AE001863; GB:AE001825; MID:g6460670; PIDN:AAF12476.1; PID:g64607  
A:Experimental source: strain R1  
C:Genetics:cccc

A:Gene: DRA0290  
A:Map position: 2  
C:Superfamily: cell division protein ftsH; FtsH/SEC18/CDC48-type ATP-binding domain hom

Query Match	44.8%;	Score 43;	DB 2;	Length 655;
Best Local Similarity	56.2%;	Pred. No. 46;		
Matches	9;	Conservative	2;	Mismatches 5; Indels 0; Gaps 0

QY 2 AFEPHGYLLTAANSPG 17  
|| | : |||| ||  
ACQ 3EPH2CYUWMAVUTDQ 47E

## RESULT 11

RESULT 11  
T20556  
Hypothetical protein F07H5.9 - *Caenorhabditis elegans*  
C|Species: *Caenorhabditis elegans*  
C|Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
C|Accession: T20556  
R|Steward, C.  
Submitted to the EMBL Data Library, December 1995  
!Reference number: Z19292

A1:Accession: T20556  
A1:Status: preliminary; translated from GB/EMBL/DBJ  
A1:Molecule type: DNA

A1:Residues: 1-452 <MIL>  
A1:Cross-references: EMBL:Z68314; P1DN:CAA92657.1; GSPDB:GN00020; CESP:F07H5.9  
A1:Experimental source: clone F07H5

A;Gene: CESP:F07H5.9  
A;Map position: 2  
A;Introns: 75/3; 171/3; 273/1; 331/3; 378/3  
C;Superfamily: mammalian acid phosphatase

Query Match	44.3%	Score 42.5;	DB 2;	Length 452;
Best Local Similarity	66.7%	Pred. No. 38;		
Matches 10;	Conservative 2;	Mismatches 2;	Indels 1;	Gaps 1

```

QY      2 AFEPHG-Y-LTTA-VS 15
      | : ||| | | : ||| |
Db      349 AIKPHGYPLYSAAVS 363

```

RESULT 12  
 GMEBO  
 gastrin precursor [validated] - bovine  
 C:/Species: Bos primigenius taurus (cattle)  
 C:/Date: 31-Dec-1991 #sequence\_revision 23-Mar-1995 #text\_change 20-Oct-2000  
 C:/Accession: S14400; A14409; B01619; A01619  
 R:/Kim, S.J.; Uhm, K.N.; Kang, Y.K.; Yoo, O.J.  
 DNA Seq. 1, 181-187, 1991

A:Title: Bovine and feline gastrin cDNA sequences and the amino acid and nucleotide sequence  
A:Reference number: S14400; MUID:92127058; PMID:1773057  
A:Accession: S14400  
A:Molecule type: mRNA  
A:Residues: 1-104 <KIM>  
A:Cross-references: EMBL:X16581; NID:G648; PIDN:CAA4598.1; PID:G649  
R:Lund, T.; Olsen, J.; Rehfeld, J.F.  
Mol. Endocrinol. 3, 1585-1588, 1989  
A:Title: Cloning and sequencing of the bovine gastrin gene.  
A:Reference number: A41409; MUID:90114160; PMID:2608050  
A:Accession: A41409  
A:Molecule type: DNA  
A:Residues: 1-11, 'L', 33-36, 'R', 38-47, 'T', 49-73, 'N', 75-80, 'G', 82-95, 'W', 97-98, 'G', 100-104  
A:Cross-references: GB:M1657; NID:G163079; PIDN:AA30537.1; PID:G163080  
A:Note: the authors translated the codon CTG for residue 32 as Ala, AAT for residue 39 as Ala  
R:Agarwal, K.L.; Beacham, J.; Bentley, P.H.; Gregory, R.A.; Kenner, G.W.; Sheppard, R.C.  
Nature 219, 614-615, 1968  
A:Title: Isolation, structure and synthesis of ovine and bovine gastrins.  
A:Reference number: A01619; MUID:68357500; PMID:5665711  
A:Accession: B01619  
A:Molecule type: Protein  
A:Residues: 76-92 <AGA>  
A:Genetics:  
A:introns: 71/1  
C:Superfamily: gastrin  
C:Keywords: amidated carboxyl end; hormone; pancreas; phosphoprotein; pyroglutamic acid  
F:1-19/Domains: signal sequence #status predicted <SIG>  
F:159-92/Product: big gastrin #status predicted <BGN>  
F:76-92/Product: gastrin #status experimental <SGN>  
F:59/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted  
F:87/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimental  
F:87/Binding site: sulfide (tyr) (covalent) (partial) #status experimental  
F:92/Modified site: amidated carboxyl end (Phe) (amide in mature form from following gly

```

sugar transport system (sugar-binding protein) BH3690 [imported] - Bacillus halodurans (
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #ext_change 15-Jun-2001
C:Accession: B84111
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Maeki, N.; Fujii, F.; Hiran
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: AB3650; MUID:20512582; PMID:11058132
A:Accession: B84111
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-420 <STO>
A:Cross-references: GB:AP001519; GB:BA000004; NID:G10176109; PIDN:BA807409.1; GSPDB:GN00
C:Genetics:
A:Gene: BH3690

Query Match          43.8%; Score 42; DB 2; Length 420;
Best Local Similarity 56.2%; Pred. No. 42;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY      1 DAFPHGYLTLTAASP 16
      ||||| : ||||
DB      106 DAFAPGLTGTAVP 121

RESULT 15
T22777
Hypochemical protein F56D5.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #ext_change 15-Oct-1999
C:Accession: T22777
R:Matthews, P.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19614
A:Accession: T22777
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-449 <WIL>
A:Cross-references: EMBL:Z69662; PIDN:CAA93502.1; GSPDB:GN00022; CESP:F56D5.3
A:Experimental source: clone F56D5
C:Genetics:
A:Gene: CESP:F56D5.3
A:Map position: 4
A:Introns: 82/3; 311/3; 404/2

Query Match          43.8%; Score 42; DB 2; Length 449;
Best Local Similarity 72.7%; Pred. No. 46;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      6 HGVLTLTAASP 16
      ||||| : ||||
DB      196 HGVLTLTAASP 206

```

Search completed: March 22, 2004, 07:01:24  
Job time : 2.17065 secs

## RESULT 14

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:31:13 ; Search time 0.690384 Seconds

(without alignments)  
1357.597 Million cell updates/sec

Title: US-09-662-293-6

Perfect score: 96

Sequence: 1 DAFEPHGYLLTPAAVSPGK 18

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59	61.5	508	1 CH11_DROME	Q9W5U3 drosophila
2	52	54.2	458	1 CH13_DROME	Q9W5U3 drosophila
3	45.5	47.4	258	1 YGBM_ECOLI	Q46891 escherichia
4	45.5	47.4	1523	1 DPOL_THERM	P74918 thermococcus
5	45.5	46.9	262	1 MOAF_KLEAE	P54796 klebsiella
6	43.5	45.3	1668	1 DPOL_THERM	Q9H005 thermococcus
7	43	44.8	88	1 XKDR_BACSU	P54337 bacillus su
8	42	43.8	104	1 GAST_BOVIN	P01352 bos taurus
9	42	43.8	255	1 RCEL_RHOPH	P51750 rhodospirill
10	42	43.8	304	1 PQGE_GLUOX	Q913b3 gluconobact
11	42	43.8	500	1 CISA_BACSU	P17867 bacillus su
12	41	42.7	275	1 RCEL_RHOPH	P10717 rhodospirill
13	41	42.7	620	1 EXON_HSV2	P06489 herpes simp
14	41	42.7	3313	1 CLR3_HUMAN	Q9N977 homo sapien
15	41	42.7	3313	1 CLR3_RAT	O88278 rattus norv
16	40	41.7	251	1 Y665_TREPA	O83671 treponema p
17	40	41.7	318	1 NK25_RAT	O35677 rattus norv
18	40	41.7	435	1 PNK1_MOUSE	O70447 mus musculu
19	40	41.7	459	1 Y819_PYRHO	O58549 pyrococcus
20	40	41.7	501	1 ARAA_MYCSM	O9rh22 mycobacteri
21	40	41.7	583	1 FOUO_DROME	P54360 drosophila
22	40	41.7	716	1 RRP2_IAMN	P21427 infuensa a
23	40	41.7	716	1 RRP2_IAMN	P21427 infuensa a
24	40	41.7	716	1 RRP2_IAMN	P21427 infuensa a
25	40	41.7	716	1 RRP2_IAMN	P21427 infuensa a
26	40	41.7	716	1 RRP2_IAMN	P21427 infuensa a
27	40	41.7	716	1 RRP2_IAMN	P21427 infuensa a
28	40	41.7	716	1 RRP2_IAMN	P21427 infuensa a
29	40	41.7	716	1 RRP2_IAMN	P21427 infuensa a
30	40	41.7	716	1 RRP2_IAMN	P21427 infuensa a
31	40	41.7	716	1 RRP2_IAMN	P21427 infuensa a
32	40	41.7	716	1 RRP2_IAMN	P21427 infuensa a
33	40	41.7	716	1 RRP2_IAMN	P21427 infuensa a

34	40	41.7	1690	1 PPOC_THERM	P36252 thermotoga
35	40	41.7	1720	1 FRSH_CHLYU	P56365 chlorella v
36	40	41.7	2464	1 NABP_MOUSE	P14873 mus musculu
37	39.5	41.1	157	1 YHHE_LACIA	Q9ch89 lactococcus
38	39	40.6	104	1 GAST_SHEEP	O02686 ovis aries
39	39	40.6	318	1 GBLP_CHLRE	P53187 chlamydomon
40	39	40.6	319	1 GBLP_YEAST	P38011 saccharomyc
41	39	40.6	327	1 Y080_BACAN	Q9rmv5 bacillus an
42	39	40.6	350	1 UNRI_HUMAN	Q9y314 homo sapien
43	39	40.6	351	1 UNRI_MOUSE	Q92122 mus musculu
44	39	40.6	356	1 HIS8_LACPL	O88ue6 lactobacill
45	39	40.6	504	1 CHIT_BRUMA	P29030 brugia mala

## ALIGNMENTS

RESULT 1  
CH11\_DROME STANDARD, PRT, 508 AA.  
AC Q9W5U3; 017420;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Probable chitinase I (BC 3.2.1.14).  
GN CH11 OR CG17682.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RX MEDLINE=22426071; PubMed=12537574;  
RA Hoskins R.A., Smith C.D., Carlson J.W., Carvalho A.B., Halpern A.,  
RA Kaminker J.S., Kennedy C., Mungall C.J., Sullivan B.A., Sutton G.G.,  
RA Yasuhara J.C., Makimoto B.T., Myers E.W., Celisner S.E., Rubin G.M.,  
RA Karpen G.H.;  
RT "Heterochromatic sequences in a Drosophila whole-genome shotgun  
RT assembly.";  
RT Genome Biol. 3:RESEARCH0085.1-RESEARCH0085.16(2002).  
[2]  
SEQUENCE OF 151-263 FROM N.A.  
RC STRAIN=Canton-S;  
RX MEDLINE=98324849; PubMed=9662472;  
RA de la Vega H., Specht C.A., Liu Y., Robbins P.W.;  
RT "Chitinases are a multi-gene family in Aedes, Anopheles and  
RT Drosophila.";  
RT Insect Mol. Biol. 7:233-239(1998).  
RL -I- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-  
RL acetyl-D-glucosamine polymers of chitin.  
CC -I- SIMILARITY: Belongs to chitinase class II (family 18 of glycosyl  
CC hydrolases).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL, AF026500; AAB81858.1; -  
CC FLYBASE; FBgn022703; Ch11  
CC InterPro; IPR001223; Glyco\_hydro\_18.  
CC InterPro; IPR001579; Glyco\_hydro\_18a5.  
CC Pfam; PF00704; Glyco\_hydro\_18; 2.  
CC ProDom; PD000471; Glyco\_hydro\_18; 2.  
CC SMART; SM00636; Glyco\_18; 1.  
CC PROSITE; PS01095; CHITINASE\_18; 1.  
KW Hydrolase; Glycosidase; Chitin degradation; Multi-gene family.  
KW ACT\_SITE 264 264 PROTON DONOR (BY SIMILARITY).  
FT

QY	SEQUENCE	508 AA;	57751 MW;	26CA23B02EFDEB97	CRC64;
Query Match		61.5%;	Score 59;	DB 1;	Length 508;
Best Local Similarity		64.7%;	Pred. No. 0.052;		
Matches	11; Conservative	3;	Mismatches	3;	Indels 0; Gaps 0;
DQ	2 AFEPHYLLTAAPSCK 18				
	: : : : :				
DQ	294 AFQPRGLTSLAAPSCK 310				
RESULT 2					
CH3_DROME	STANDARD;	PRT;	458 AA.		
ID	CH3_DROME				
AC	Q9M5U2; O17422;				
DT	10-OCT-2003 (Rel. 42, Created)				
DT	10-OCT-2003 (Rel. 42, Last sequence update)				
DE	15-MAR-2004 (Rel. 43, Last annotation update)				
DR	Probable chitinase 3 (EC 3.2.1.14).				
GN	CHT3 OR CG18140.				
OS	Drosophila melanogaster (Fruit fly).				
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;				
CC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
CC	Ephydroidea; Drosophilidae; Drosophila.				
CC	NCBI_TaxID=7227;				
RN	SEQUENCE FROM N.A.				
RP	STRAIN=Berkely;				
RX	MEDLINE=22426071; PubMed=12537574;				
RA	Hoskins R.A., Smith C.D., Carlson J.W., Carvalho A.B., Halpern A.,				
RA	Kantink R.J.S., Kennedy C., Mungall C.J., Sullivan B.A., Sutton G.G.,				
RA	Yasuhara J.C., Wakimoto B.T., Myers E.W., Celniker S.E., Rubin G.M.,				
RA	Katpen G.H.;				
RT	"Heterochromatic sequences in a Drosophila whole-genome shotgun				
RT	assembly.";				
RL	Genome Biol. 3:RESEARCH0085.1-RESEARCH0085.16(2002).				
RN	[2]				
RP	SEQUENCE OF 182-294 FROM N.A.				
RC	STRAIN=Cancon-S;				
RX	MEDLINE=98324849; PubMed=9662472;				
RA	de la Vega H., Specht C.A., Liu Y., Robbins P.W.;				
RT	"Chitinases are a multi-gene family in Aedes, Anopheles and				
RT	Drosophila.";				
RL	Insect Mol. Biol. 7:233-239(1998).				
CC	-1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-				
CC	acetyl-D-glucosamine polymers of chitin.				
CC	-1- SIMILARITY: Belongs to chitinase class II (family 18 of glycosyl				
CC	hydrolases).				
CC	-1- SIMILARITY: Contains 2 chitin-binding type-2 domains.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sb.ch/announce">http://www.isb-sb.ch/announce</a>				
CC	or send an email to <a href="mailto:license@isb-sb.ch">license@isb-sb.ch</a> ).				
CC	-----				
DR	EMBL; AF026503; AAB81860.1; -				
DR	FLYBASE; FBgn0022701; Cht3.				
DR	InterPro; IPR002557; Chitin_bind_Pera.				
DR	InterPro; IPR001223; Glyco_hydro_18.				
DR	InterPro; IPR001579; Glyco_hydro_18A.				
DR	Pfam; PF01607; CEM_14; 2.				
DR	Pfam; PF00704; Glyco_hydro_18; 1.				
DR	ProDom; PD000471; Glyco_hydro_18; 1.				
DR	SMART; SM00494; ChtBD2; 2.				
DR	SMART; SM00636; Glyco_18; 1.				
DR	PROSITE; PS00940; CHIT BIND II; 2.				
DR	PROSITE; PS01095; CHITINASE_18; FALSE NEG.				
KW	Hydrolase; Glycosidase; Chitin degradation; Chitin-binding;				
KW	Multi-gene family; Repeat.				
FT	DOMAIN	5	58	CHITIN-BINDING TYPE-2 1.	

```

FT DOMAIN 74 128 CHITIN-BINDING TYPE-2.2.
PT ACT SITE 295 295 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 458 AA; 52330 MW; 4A063190B7926248 CRC64;

Query Match 54.2%; Score 52; DB 1; Length 458;
Best Local Similarity 50.0%; Pred. No. 0.66;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 1 DAFEPHG-YLITAAVSPGK 18
: ||| : ||| :
Db 324 EAFPRGLMSTAVSPSR 341

RESULT 3
ID YGBM_ECOLI STANDARD; PRT; 258 AA.
AC Q46891;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein ygbM.
DN YGBM OR B2739.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
CX NCBI_TaxID=562;
CK [1]
RN SEQUENCE FROM N.A.
RP STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Bluttner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -1- SIMILARITY: BELONGS TO THE HI1 FAMILY. STRONG, TO
CC H. INFLUENZAE HI1013.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/annouce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U29579; AAA69249.1; -.
DR EMBL; AE000357; AAC75781.1; -.
DR PIR; G65054; G65054.
DR PDB; 1K77; 30-OCT-02.
DR EcoGene; EG13107; ygbM.
DR InterPro; IPR001719; AP_endonuclease2.
DR Pfam; PF01261; AP_endonc_2; 1.
DR Hypothetical protein; Complete proteome; 3D-structure.
SQ SEQUENCE 258 AA; 29217 MW; 07392B724336D7DD CRC64;

Query Match 47.4%; Score 45.5; DB 1; Length 258;
Best Local Similarity 55.6%; Pred. No. 4.3;
Matches 10; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

Qy 1 DAFEPHG-YLITAAVSPG 17
: ||| : ||| :
Db 131 DRFAPHGRIIVLSPG 148

RESULT 4
DPOL_THEFM STANDARD; PRT; 1523 AA.
AC P74918;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)

```

DE DNA polymerase (EC 2.7.7.7) (Pol Tfu) [Contains: Endonuclease PI-Tfu  
 DE (EC 3.1.1.1) (Tfu pol-1 intein); Endonuclease PI-TfuIT (EC 3.1.1.1)]  
 DE (Tfu pol-2 intein)].  
 GN POL.  
 OS Thermococcus funiculans.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Thermococcus.  
 OC NCBI\_TaxID=46540;  
 RX MEDLINE=20112788; PubMed=10644683;  
 RA Saves I., O'Garra V., Dietrich J., Nasson J.-M.;  
 RT "Inteins of Thermococcus funiculans DNA polymerase are endonucleases  
 RT with distinct enzymatic behaviors.";  
 RL J. Biol. Chem. 275:2335-2341(2000).  
 CC -1- FUNCTION: PI-TfuIT recognizes and cleaves a minimal sequence of 16  
 CC base pairs (bp) on supercoiled DNA with either Mg(2+) or Mn(2+) as  
 CC cofactor. It cleaves linear DNA only with Mg(2+) and requires a  
 CC 19-bp minimal recognition sequence. The optimal temperature for  
 CC activity is 70 degrees Celsius.  
 CC -1- FUNCTION: PI-TfuIT is a highly active homing endonuclease using  
 CC Mg(2+) as cofactor. Its minimal recognition and cleavage site is  
 CC 21 bp long either on linear or circular DNA substrates. Its  
 CC endonuclease activity is strongly inhibited by the 3' digestion  
 CC product, which remains bound to the enzyme after the cleavage  
 CC reaction. The optimal temperature for activity is 70 degrees  
 CC Celsius.  
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
 CC + {DNA}(n).  
 CC -1- PTM: This protein undergoes a protein self splicing that involves  
 CC a post-translational excision of the two intervening regions  
 CC (inteins) followed by peptide ligation.  
 CC -1- SIMILARITY: Belongs to the DNA polymerase type-B family.  
 CC -1- SIMILARITY: In the intein section; belongs to the homing  
 CC endonuclease family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z69882; CA93738.1; -.  
 DR HSSP; P56689; ITGO.  
 DR REBASE; 4500; PI-TfuIT.  
 DR REBASE; 4501; PI-TfuIT.  
 DR InterPro; IPR006172; DNA\_pol\_B.  
 DR InterPro; IPR006134; DNA\_pol\_B\_dom.  
 DR InterPro; IPR006133; DNA\_pol\_B\_exo.  
 DR InterPro; IPR003587; Hedgehog\_hintc.  
 DR InterPro; IPR003586; Hedgehog\_hintc.  
 DR InterPro; IPR006142; INTEIN.  
 DR InterPro; IPR004042; intein\_endonuc.  
 DR InterPro; IPR006141; intein\_s.  
 DR Pfam; PF00136; DNA\_pol\_B; 2.  
 DR Pfam; PF03104; DNA\_pol\_B\_exo; 1.  
 DR PRINTS; PR00379; INTEIN.  
 DR SMART; SM00305; Hintc; 2.  
 DR SMART; SM00306; Hintn; 2.  
 DR SMART; SM00486; POLBc; 1.  
 DR TIGRFAMs; TIGR01443; intein\_Cterm; 2.  
 DR TIGRFAMs; TIGR01445; intein\_Nterm; 2.  
 DR TIGRFAMs; TIGR00592; pol2; 1.  
 DR PROSITE; PS00116; DNA\_POLYMERASE\_B; FALSE\_NEG.

DR PROSITE; PS00819; INTEIN\_C\_TER; 2.  
 DR PROSITE; PS00819; INTEIN\_ENDONUCLEASE; 2.  
 DR PROSITE; PS00817; INTEIN\_N\_TER; 2.  
 KW Transferase; DNA-directed DNA polymerase; DNA replication;  
 KW DNA-binding; Hydrolase; Endonuclease; Autocatalytic cleavage;  
 KW Protein splicing; Intron homing; Magnesium; Manganese;  
 FT CHAIN 1 406  
 FT CHAIN 407 766  
 FT CHAIN 767 900  
 FT CHAIN 901 1282  
 FT CHAIN 1283 1523  
 SQ SEQUENCE 1523 AA; 175917 MW; 7A2AC8236B2E5F5 CRC64;  
 Query Match 47.4%; Score 45.5; DB 1; Length 1523;  
 Best Local Similarity 52.4%; Pred. No. 26;  
 Matches 11; Conservative 1; Mismatches 6; Indels 3; Gaps 1;  
 QY 1 DAFEPH---GYLTPAASP 18  
 DB 983 DVTEDHSLIGYLTSTKVPK 1003  
 RESULT 5  
 MOAF\_KLEAF STANDARD; PRT; 262 AA.  
 AC P54796;  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE MOAF protein precursor.  
 OS MOAF.  
 GN Klebsiella aerogenes.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Klebsiella.  
 OX NCBI\_TaxID=28451;  
 RP [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 14-21.  
 RX STRAIN=WT0;  
 RX MEDLINE=96060844; PubMed=7590328;  
 RA Azakami H., Sugino H., Iwata N., Yokoro N., Yamashita M., Hirooka Y.;  
 RT "A Klebsiella aerogenes moaf operon is controlled by the positive  
 RT Moaf regulator of the monamine regulon.";  
 RL Gene 164:89-94 (1995).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D28533; BAA05890.1; -.  
 DR REBASE; 14; MOAF\_PROTEIN.  
 DR PROPEP 1 13  
 FT CHAIN 14 262  
 SQ SEQUENCE 262 AA; 29157 MW; 0AE3F94967A2A2A8 CRC64;  
 Query Match 46.9%; Score 45; DB 1; Length 262;  
 Best Local Similarity 56.2%; Pred. No. 53;  
 Matches 9; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 1 DAFEPHGYLTPAASP 16  
 DB 15 DGFAPHGMLTAVSLP 30  
 RESULT 6  
 DPOL\_THERY STANDARD; PRT; 1668 AA.  
 ID DPOL\_THERY  
 AC Q9HH05;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE DNA polymerase (EC 2.7.7.7) [Contains: Endonuclease P1-Thy1I  
(EC 3.1.1.1) (Thy pol-1 intein); Endonuclease P1-Thy1 (EC 3.1.1.1) (Thy pol-2 intein)] (Fragment).  
DE (Thy pol-2 intein) (Fragment).  
GN POL.  
OS Thermococcus hydrothermalis.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Thermococcus.  
OX NCBI\_TaxID=46539;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Querellou J.-J.E., Camdon M.A., Lesongeur F.O., Barbier G.,  
RT "Thermococcales taxonomy and phylogeny based on the comparative use of  
RT 16S rDNA, 16S-23S rDNA intergenic spacer and family B DNA polymerase  
RT genes";  
RT Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP CHARACTERIZATION OF P1-THY1.  
RX MEDLINE=20512590; PubMed=11058140;  
RA Saves I., Eleaume H., Dietrich J., Masson J.-M.,  
RT "The Thy pol-2 intein of Thermococcus hydrothermalis is an  
RT isoschizomer of P1-T111 and P1-T111 endonucleases.",  
RL Nucleic Acids Res. 28:4391-4396(2000).  
CC -1- FUNCTION: In addition to polymerase activity, this DNA polymerase  
CC exhibits 3' to 5' exonuclease activity (by similarity).  
CC -1- FUNCTION: P1-THY1 AND P1-THY1 ARE ENDONUCLEASES. P1-THY1 CLEAVES  
CC THE INTERINTEIN SEQUENCE OF THE THY DNA POL GENE. IT REQUIRES A 21-  
CC BP MINIMAL RECOGNITION SEQUENCE.  
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
CC + {DNA} (N).  
CC -1- PFM: This protein undergoes a protein self splicing that involves  
CC a post-translational excision of the intervening region (intein)  
CC followed by peptide ligation (Potential).  
CC -1- SIMILARITY: Belongs to the DNA polymerase type-B family.  
CC -1- SIMILARITY: In the intein section, belongs to the homing  
CC endonuclease family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AJ245819; CAC18555.1; .  
DR HSBP; P56689; IRGO.  
DR REBASE; 4832; P1-THY1.  
DR InterPro; IPR006172; DNA\_pol\_B.  
DR InterPro; IPR006134; DNA\_pol\_B\_dom.  
DR InterPro; IPR006133; DNA\_pol\_B\_exo.  
DR InterPro; IPR003587; Hedgehog\_hint\_N.  
DR InterPro; IPR003586; Hedgehog\_hintC.  
DR InterPro; IPR006142; INTEIN.  
DR InterPro; IPR004042; Intein\_endonuc.  
DR InterPro; IPR006141; Intein\_S.  
DR InterPro; IPR004578; Pol2.  
DR Pfam; PF00136; DNA\_pol\_B; 3.  
DR Pfam; PF0104; DNA\_pol\_B\_exo; 1.  
DR PRINTS; PR00379; INTEIN.  
DR SMART; SM00305; HincC; 2.  
DR SMART; SM00306; HincN; 1.  
DR SMART; SM00486; POLBC; 1.  
DR TIGRFAMs; TIGR01443; intein\_Cterm; 2.  
DR TIGRFAMs; TIGR01445; intein\_Nterm; 2.  
DR TIGRFAMs; TIGR00592; pol2; 2.  
DR PROSITE; PS00116; DNA\_POLYMERASE\_B; FALSE\_NEG.  
DR PROSITE; PS50818; INTEIN\_C\_TER; 2.  
DR PROSITE; PS50819; INTEIN\_ENDONUCLEASE; 2.  
DR PROSITE; PS50817; INTEIN\_N\_TER; 2.  
KM Transferrase; DNA-directed DNA polymerase; DNA replication;  
KM DNA-binding; Hydrolase; Nuclease; Exonuclease; Multifunctional enzyme;  
KM Protein splicing; Autocatalytic cleavage; Endonuclease; Intronic homing.  
FT 1

FT CHAIN 1 458 DNA POLYMERASE, 1ST PART.  
FT CHAIN 459 995 ENDONUCLEASE P1-THY1 (POTENTIAL).  
FT CHAIN 996 1044 DNA POLYMERASE, 2ND PART.  
FT CHAIN 1045 1433 ENDONUCLEASE P1-THY1 (POTENTIAL).  
FT CHAIN 1434 1668 DNA POLYMERASE, 3RD PART.  
SQ SEQUENCE 1668 AA; 193319 MW; 5EE805FEDFA1C8 CRC64;  
  
Query Match 45.3%; Score 43.5; DB 1; Length 1668;  
Best Local Similarity 47.6%; Pred. No. 60;  
Matches 10; Conservative 2; Mismatches 6; Indels 3; Gaps 1;  
  
QY 1 DAFEPH---GTLTAUVPFGK 18  
DB 1127 DVTEDHSLIGYMTSKVXFGK 1147  
  
RESULT 7  
XKDR BACSU STANDARD; PRT; 88 AA.  
ID XKDR BACSU  
AC P54337;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Phage-like element PBX protein XKDR.  
DE XKDR OR BSU12710.  
GN Bacillus subtilis.  
OS Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA Krogh S., O'Reilly M., Nolan N., Devine K.M.;  
RL Submitted (Mar-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunat P., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Barteto M.G., Bessieres P., Bortolin A., Borchert S.,  
RA Bortolin R., Bortolin L., Brans A., Braun M., Bignelli S.C., Bron S.,  
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,  
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
RA Entian K.D., Ertington J., Fabre C., Ferrati E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
RA Guisepi G., Guy B.J., Haga K., Haeck J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Maubert C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noack M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudea B., Park S.H.,  
RA Parro V., Pohl T.M., Portecelle D., Portocarrero S., Prescott A.M.,  
RA Presecan E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,  
RA Sato T., Scanlan E., Schleich S., Schneider P., Scofield P.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
RA Toso V., Uchiyama S., Vandenbol M., Vannier F., Vassaretto A.,  
RA Viari A., Wambutt R., Wedler E., Wedler H., Westenberger T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis".  
RL Nature 390:249-256(1997).  
CC -1- SIMILARITY: STRONG, TO B. SUBTILIS YOB8.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its



use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; 270177; CAA94039.1; -

DR EMBL; 299110; CAB13128.1; -

DR PIR; A69733; A69733.

DR Subtilisin; BG11552; XKDR.

KM Complete proteome.

SQ SEQUENCE 88 AA; 9381 MW; 5C814BA818CB847 CRC64;

Query Match 44.8%; Score 43; DB 1; Length 88;  
Best Local Similarity 50.0%; Pred. No. 3.8;  
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 DAFEPHYLTITAAVSPGK 18  
DB 63 DALEPGRLMTALTGQ 80

RESULT 8  
GAST\_BOVIN STANDARD; PRT; 104 AA.

AC P01352; Q28114;  
DT 21-UTR-1986 (Rel. 01, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Gastrin precursor.  
GN GAS.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=90114160; PubMed=2608050;  
RA Lund T., Olsen J., Rehfeld J.F.;  
RT "Cloning and sequencing of the bovine gastrin gene.";  
RL Mol. Endocrinol. 3:1585-1588(1989).  
[2]  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92127058; PubMed=1773057;  
RA Kim S.J., Um K.N., Kang Y.K., Yoo O.J.;  
RT "Bovine and feline gastrin cDNA sequences and the amino acid and  
nucleotide sequence homologies among mammalian species.";  
RL DNA Seq. 1:181-187(1991).  
[3]  
RN [3]  
RP SEQUENCE OF 76-92.  
RX MEDLINE=68357500; PubMed=565711;  
RA Aharwal K.L., Beacham J., Bentley P.H., Gregory R.A., Kemner G.W.,  
RA Sheppard R.C., Tracy H.J.;  
RT "Isolation, structure and synthesis of ovine and bovine gastrins.";  
RL Nature 219:614-615(1968).  
CC -1- FUNCTION: Gastrin stimulates the stomach mucosa to produce and  
secrete hydrochloric acid and the pancreas to secrete its  
digestive enzymes. It also stimulates smooth muscle contraction  
and increases blood circulation and water secretion in the stomach  
and intestine.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- SIMILARITY: Belongs to the gastrin/cholecystokinin family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X16581; CAA34598.1; -

DR PIR; S14400; GMB0.

DR InterPro; IPR001651; Gastrin.

DR Pfam; PF00918; Gastrin; 1.

DR SMART; SM00029; GASTRIN; 1.

DR PROSITE; PS00259; GASTRIN; 1.

KM Hormone; cleavage on pair of basic residues; Amidation; Sulfation;  
KW Signal; phosphorylation; Pyrrolidone carboxylic acid.  
FT SIGNAL 1 21  
FT PEPTIDE 59 92 BIG GASTRIN (GASTRIN 34).  
FT PEPTIDE 76 92 GASTRIN.  
FT MOD\_RES 59 59 PYRROLIDONE CARBOXYLIC ACID.  
FT MOD\_RES 76 76 PYRROLIDONE CARBOXYLIC ACID.  
FT MOD\_RES 87 87 SULFATION (PARTIAL).  
FT MOD\_RES 92 92 AMIDATION (G-93 PROVIDE AMIDE GROUP).  
FT MOD\_RES 96 96 PHOSPHORYLATION (BY SIMILARITY).  
FT CONFLICT 32 32 A -> L (IN REF. 1).  
FT CONFLICT 37 37 G -> R (IN REF. 1).  
FT CONFLICT 48 48 N -> T (IN REF. 1).  
FT CONFLICT 74 74 K -> N (IN REF. 1).  
FT CONFLICT 81 81 E -> G (IN REF. 1).  
FT CONFLICT 96 96 S -> M (IN REF. 1).  
FT CONFLICT 99 99 E -> G (IN REF. 1).  
SQ SEQUENCE 104 AA; 11573 MW; 54D03BF200F299F2 CRC64;

Query Match 43.8%; Score 42; DB 1; Length 104;  
Best Local Similarity 43.8%; Pred. No. 6.5;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 AFEPHYLTITAAVSPG 17  
DB 22 SWKPSHLQDAPVAPG 37

RESULT 9  
RCEL\_RHOPH STANDARD; PRT; 255 AA.

AC P51750;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Reaction center protein L chain (Photosynthetic reaction center L  
subunit) (Fragment).  
GN PUF.L  
OS Rhodospirillum rubrum.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;  
OC Rhodospirillaceae; Rhodospirillum.  
OX NCBI\_TaxID=1084;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Nagashima K.V., Hiraiishi A., Shimada K., Matsura K.;  
RL Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.  
CC -1- FUNCTION: THE REACTION CENTER IS A MEMBRANE-BOUND COMPLEX THAT  
CC MEDIATES THE INITIAL PHOTOCHEMICAL EVENT IN THE ELECTRON TRANSFER  
CC PROCESS OF PHOTOSYNTHESIS.  
CC -1- SUBUNIT: REACTION CENTER IS COMPOSED OF FOUR BACTERIOCHLOROPHYLLS,  
CC TWO BACTERIOPEOPHYTINS, TWO UBIQUINONES, ONE IRON, AND THREE  
CC HIGHLY HYDROPHOBIC POLYPEPTIDE CHAINS (DESIGNATED L, M, AND H).  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- SIMILARITY: BELONGS TO THE REACTION CENTER L/M CHAINS / PSBA /  
CC PSBD FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; D50681; BAA09329.1; -

DR HSSP; P06009; 6PRC.

DR InterPro; IPR005871; Photo\_L.

DR InterPro; IPR000484; Photo\_RC.  
 DR Pfam; PF00124; photoRC; 1.  
 DR PRINTS; PR00256; REACTCENTRE.  
 DR PRODOM; PD000551; Photo RC; 1.  
 DR TIGRFAMs; TIGR01157; PUF; 1.  
 DR PROSITE; PS00244; REACTION\_CENTER; 1.  
 KW Electon transport; Photosynthesis; Reaction center;  
 Bacteriochlorophyll; Iron; Magnesium; Transmembrane.  
 FT TRANSMEM 12 35 POTENTIAL.  
 FT TRANSMEM 64 92 POTENTIAL.  
 FT TRANSMEM 95 120 POTENTIAL.  
 FT TRANSMEM 150 179 POTENTIAL.  
 FT TRANSMEM 205 231 POTENTIAL.  
 FT METAL 133 133 MAGNESIUM (BACTERIOCHLOROPHYLL B AXIAL  
 LIGAND) (BY SIMILARITY).  
 FT METAL 153 153 MAGNESIUM (BACTERIOCHLOROPHYLL B AXIAL  
 LIGAND) (BY SIMILARITY).  
 FT METAL 170 170 IRON (NON HEME) (BY SIMILARITY).  
 FT METAL 210 210 IRON (NON HEME) (BY SIMILARITY).  
 FT BINDING 196 196 QUINONE B (BY SIMILARITY).  
 SQ SEQUENCE 255 AA; 28489 MW; 22DD81DDCFE54DC4 CRC64;  
 Query Match 43.8%; Score 42; DB 1; Length 255;  
 Best Local Similarity 41.2%; Pred. No. 16;  
 Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AEPHGVLTAAVSPGK 18  
 Db 166 AFGMHGSIILSVLPNGK 182

RESULT 10  
 PQOB\_GLUOX STANDARD; PRT; 304 AA.  
 AC 0913B3;  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Coenzyme PQQ synthetase protein B (Pyroloquinoline quinone  
 biosynthesis protein B).  
 GN PQOB.  
 OS Gluconobacter oxydans (Gluconobacter suboxydans).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;  
 OC Acetobacteraceae; Gluconobacter.  
 OX NCBI\_TaxID=442;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 9937;  
 RX MEDLINE=20564161; PubMed=11111029;  
 RA Felder M., Gupta A., Varma V., Kumar A., Qazi G.N., Cullum J.;  
 RT "The pyroloquinoline quinone synthetase genes of Gluconobacter  
 oxydans";  
 RU FEMS Microbiol. Lett. 193:231-236(2000).  
 CC -1- FUNCTION: May be involved in the transport of PQQ or its precursor  
 to the periplasm (By similarity).  
 CC -1- PATHWAY: Pyroloquinoline quinone (PQQ) biosynthesis.  
 CC -1- SIMILARITY: Belongs to the pqbB family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; AJ277117; CAB8319.1; -  
 DR HAMAP; MF 00653; -; 1.  
 KW PQQ biosynthesis; Transport.  
 SQ SEQUENCE 304 AA; 32323 MW; E9C54F778A8E2732 CRC64;

Query Match 43.8%; Score 42; DB 1; Length 304;

Best Local Similarity 64.3%; Pred. No. 19;  
 Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 5 PEGVLLTAAVSPGK 18  
 Db 152 PEGVLLTAAVSPGK 165

RESULT 11  
 CISA\_BACSU STANDARD; PRT; 500 AA.  
 AC P17867;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Putative DNA recombinase.  
 GN CISA OR SPOIVCA OR BSU25770.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90130285; PubMed=2105293;  
 RA Sato T., Samori Y., Kobayashi Y.;  
 RT "The cisa cistron of Bacillus subtilis sporulation gene spoIVC  
 encodes a protein homologous to a site-specific recombinase";  
 RU J. Bacteriol. 172:1092-1098(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / JH642;  
 RX MEDLINE=95219086; PubMed=7704261;  
 RA Takemaru K.-I., Mizuno M., Sato T., Takeuchi M., Kobayashi Y.;  
 RT "Complete nucleotide sequence of a skin element excised by DNA  
 rearrangement during sporulation in Bacillus subtilis";  
 RU Microbiology 141:323-327(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / JH642;  
 RX MEDLINE=97124195; PubMed=8969508;  
 RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,  
 Kobayashi Y.;  
 RT "Systematic sequencing of the 283 kb 210 degrees-212 degrees region of  
 the Bacillus subtilis genome containing the skin element and many  
 sporulation genes";  
 RU Microbiology 142:3103-3111(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 Azevedo V., Bertolo M.G., Bessieres P., Bolotin A., Borchert S.,  
 Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
 Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,  
 Denicof F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 Entian K.D., Errington J., Fabret C., Ferrari B., Fougere D.,  
 Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,  
 Gilm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 Giuseppe G., Guy B.J., Haga K., Hachez J., Harwood C.R., Henaut A.,  
 Hilbert H., Holstappel S., Hosono S., Hullo M.F., Iraya M., Jones L.,  
 Uoris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 Kurita K., Lapidus A., Lardinois S., Jauber J., Lazarevic V.,  
 Lee S.M., Levine A., Liu H., Maesuda S., Mauel C., Medigue C.,  
 Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 Patro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 Pressacq E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,  
 Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,  
 Seiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
 Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,  
 Takeuchi M., Tamakoshi A., Tanaka T., Terpetta P., Tognoni A.,



```

RA MEDLINE=86144016; PubMed=3005509;
RX Draper K.G., Devi-Rao G., Cosca R.H., Blair E.D., Thompson R.L.,
RA Wagner E.K.,
RT "Characterization of the genes encoding herpes simplex virus type 1
RL J. Virol. 57:1023-1036(1986).
CC -1- SIMILARITY: Belongs to the herpesviruses alkaline exonuclease
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC -----
DR EMBL; M1854; AAA45834.1; -;
DR EMBL; M1854; AAA45835.1; ALT INTN.
DR InterPro; IPR001616; Herpes_alik_exo.
DR Pfam; PF01771; Herpes_alik_exo_1.
DR PRINTS; PRO0924; ALKEXNUCLASE.
DR Hydrolase; Nuclease; Exonuclease.
KW SEQUENCE 620 AA; 66199 MW; 3E4E89AC766414B7 CRC64;

Query Match 42.7%; Score 41; DB 1; Length 620;
Best Local Similarity 53.8%; Pred. NO. 57;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Cy 4 EPHGLTAAVSP 16
Db 337 DPHGLTAAVPGTP 349

RESULT 14
CLR3_HUMAN STANDARD; PRT; 3312 AA.
ID CLR3_HUMAN
AC Q9NTQ7; 075092;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cadherin EGF LAG seven-pass G-type receptor 3 precursor (Flamingo
DE homolog 1) (hml1) (Multiple epidermal growth factor-like domains 2)
DE (Epidermal growth factor-like 1).
DE CELSR3 OR CDHF11 OR PM11 OR EGFL1 OR MEGF2.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RX [1]
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20202599; PubMed=10716726;
RA Mu O., Maniatis T.;
RT "Large exons encoding multiple ecdodomains are a characteristic
RT feature of protocadherin genes."
RL Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).
CC [2]
CC SEQUENCE OF 1954-3312 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98360089; PubMed=9693030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-trap screening."
RL Genomics 51:27-34(1998).
CC -1- FUNCTION: Receptor that may have an important role in cell/cell
CC signaling during nervous system formation.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: Belongs to family 2 of G-protein coupled receptors.
CC -1- SIMILARITY: Contains 9 cadherin domains.
CC -1- SIMILARITY: Contains 8 EGF-like domains.
CC -1- SIMILARITY: Contains 2 laminin G-like domains.
CC -1- SIMILARITY: Contains 1 laminin EGF-like domain.
CC -1- SIMILARITY: Contains 1 GPS domain.

```

CC		CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	CC
DR	EMBL; AF231023; AAF61928.1; -.		CC
DR	EMBL; AB011536; BAA32464.1; -.		CC
DR	HSSP; P00740; IEDM.		CC
DR	Genew; HGNC:3230; CELSR3.		CC
DR	MIM; 604264; -.		CC
DR	GO; GO:0005198; F.structural molecule activity; NAS.		CC
DR	InterPro; IPRO00152; Asx_hydroxyl_5.		CC
DR	InterPro; IPRO02126; Cadherin.		CC
DR	InterPro; IPRO08985; Conn_like_lig_g1.		CC
DR	InterPro; IPRO00742; EGF_2.		CC
DR	InterPro; IPRO01881; EGF_Ca.		CC
DR	InterPro; IPRO06209; EGF_like.		CC
DR	InterPro; IPRO00832; GPCR_secretin.		CC
DR	InterPro; IPRO01879; hormn_receptor.		CC
DR	InterPro; IPRO06210; IRGF.		CC
DR	InterPro; IPRO02049; Laminin_EGF.		CC
DR	InterPro; IPRO01791; Laminin_G.		CC
DR	InterPro; IPRO00203; PKD_cys_rich.		CC
DR	Pfam; PF00002; Yem_2; 1.		CC
DR	Pfam; PF00028; cadherin; 9.		CC
DR	Pfam; PF00008; EGF; 5.		CC
DR	Pfam; PF01825; GPS; 1.		CC
DR	Pfam; PF02793; HRN; 1.		CC
DR	Pfam; PF00054; laminin_G; 2.		CC
DR	PRINTS; PRO0205; CADHERIN.		CC
DR	PRINTS; PRO0011; EGFLAMININ.		CC
DR	PRINTS; PRO0249; GPCRCSECRETIN.		CC
DR	SMART; SM00112; CA; 9.		CC
DR	SMART; SM00181; EGF; 6.		CC
DR	SMART; SM00303; GPS; 1.		CC
DR	SMART; SM00008; Hormr; 1.		CC
DR	SMART; SM00282; Lamg; 2.		CC
DR	PROSITE; PS00010; ASX_HYDROXYL; 1.		CC
DR	PROSITE; PS00232; CADHERIN_1; 7.		CC
DR	PROSITE; PS00268; CADHERIN_2; 8.		CC
DR	PROSITE; PS00022; EGF_1; 6.		CC
DR	PROSITE; PS01186; EGF_2; 4.		CC
DR	PROSITE; PS00026; EGF_3; 6.		CC
DR	PROSITE; PS00649; G_PROTEIN_REC_P2_1; FALSE_NEG.		CC
DR	PROSITE; PS00650; G_PROTEIN_REC_P2_2; FALSE_NEG.		CC
DR	PROSITE; PS00227; G_PROTEIN_REC_P2_3; 1.		CC
DR	PROSITE; PS00261; G_PROTEIN_REC_P2_4; 1.		CC
DR	PROSITE; PS50221; GPS; 1.		CC
DR	PROSITE; PS50025; LAM_G_DOMAIN; 2.		CC
DR	PROSITE; PS01248; LAMININ_TYPE_EGF; 1.		CC
KW	Egf-protein coupled receptor; Transmembrane; Glycoprotein;		CC
KW	Egf-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;		CC
KW	developmental protein; Hydroxylation; Signal.		CC
FT	SIGNAL	1 32	CC
FT	CHAIN	33 3312	CC
FT			CC
FT	DOMAIN	33 2540	CC
FT	TRANSSEM	2541 2561	CC
FT	DOMAIN	2562 2572	CC
FT	TRANSSEM	2573 2593	CC
FT	DOMAIN	2594 2601	CC
FT	TRANSSEM	2602 2622	CC
FT	DOMAIN	2623 2643	CC
FT	TRANSSEM	2644 2664	CC
FT	DOMAIN	2665 2681	CC
FT	TRANSSEM	2682 2702	CC
FT	DOMAIN	2703 2725	CC
FT	TRANSSEM	2726 2746	CC
FT	DOMAIN	2747 2753	CC
FT			CC
FT	EXTRACELLULAR	(POTENTIAL).	CC
FT	RECEPTOR_3		CC
FT	EXTRACELLULAR	(POTENTIAL).	CC
FT	CYTOLASMIC	(POTENTIAL).	CC
FT	2 (POTENTIAL)		CC
FT	EXTRACELLULAR	(POTENTIAL).	CC
FT	3 (POTENTIAL)		CC
FT	CYTOLASMIC	(POTENTIAL).	CC
FT	4 (POTENTIAL)		CC
FT	EXTRACELLULAR	(POTENTIAL).	CC
FT	5 (POTENTIAL)		CC
FT	CYTOLASMIC	(POTENTIAL).	CC
FT	6 (POTENTIAL)		CC
FT	EXTRACELLULAR	(POTENTIAL).	CC

Query Match	42.7%	Score 41	DB 1	Length 3312
Best local Similarity	43.8%	Pred. No. 3.1e+02		
Matches	7	Conservative	5	Mismatches 4
				Indels 0
				Gaps 0
Qy	1	DAFEPHYLLTAAVSP	16	
		:    :		
Db	2355	DAMDPTHVLPSP	2370	

```

RESULT 15
CLR3_RAT ID CLR3_RAT STANDARD; PRT; 3313 AA.
AC 088278;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cadherin EGF LAG seven-pass G-type receptor 3 precursor (Multiple
DE epidermal growth factor-like domains 2).
GN CELSR3 OR MEGF2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
EX MEDLINE=98360089; PubMed=9693030;
FA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.,
RT "Identification of high-molecular-weight proteins with multiple
RT EGF-like motifs by motif-crash screening.";
RL Genomics 51:27-34(1998).
CC -1- FUNCTION: Receptor that may have an important role in cell/cell
CC signaling during nervous system formation.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: Expressed in the brain. Expressed in
CC cerebellum, olfactory bulb, cerebral cortex, hippocampus and
CC brain stem.
CC -1- SIMILARITY: Belongs to family 2 of G-protein coupled receptors.
CC -1- SIMILARITY: Contains 9 cadherin domains.
CC -1- SIMILARITY: Contains 8 EGF-like domains.
CC -1- SIMILARITY: Contains 2 laminin G-like domains.
CC -1- SIMILARITY: Contains 1 laminin EGF-like domain.
CC -1- SIMILARITY: Contains 1 GPS domain.
CC -----
CC This Swiss-Prot entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/).
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB011528; BAA32459.1; -.
DR HSSP; P00740; IEDM.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR008985; Cona_like_lac_gl.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR000832; GPCR_secretin.
DR InterPro; IPR006179; hormn_receptor.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR00203; PKD_cys_rich.
DR Pfam; PF00002; Tm_2; 1.
DR Pfam; PF00028; cadherin; 9.
DR Pfam; PF00008; EGF; 5.
DR Pfam; PF01825; GPS; 1.
DR Pfam; PF02793; HRM; 1.
DR Pfam; PF00054; laminin_G; 1.
DR PRINTS; PFO0205; CADHERIN.
DR PRINTS; PFO0011; EGF_LAMININ.
DR PRINTS; PFO0249; GPCRSECRETIN.
DR SMART; SM00112; CA; 9.
DR SMART; SM00181; EGF; 6.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00008; HORMR; 1.
DR SMART; SM00282; LamG; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:39:53 ; Search time 3.62201 Seconds  
(without alignments)  
1568.003 Million cell updates/sec

Title: US-09-662-293-6

Perfect score: 96

Sequence: 1 DAFEPHGYLLTAAPSFGK 18

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: sp archaea:\*  
2: sp bacteria:\*  
3: sp fungi:\*  
4: sp human:\*  
5: sp invertebrate:\*  
6: sp mammal:\*  
7: sp mhc:\*  
8: sp organelle:\*  
9: sp phage:\*  
10: sp plant:\*  
11: sp rodent:\*  
12: sp virus:\*  
13: sp vertebrate:\*  
14: sp unclassified:\*  
15: sp xvirus:\*  
16: sp bacteriaph:\*  
17: sp archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	96	100.0	555	5 Q9UER7	Q9UER7 dermatophag
2	67	69.8	2838	5 Q8MP05	Q8MP05 tenebrio mo
3	66	68.8	431	5 Q8ISHS	Q8ISHS araneus ven
4	63	65.6	527	5 P91773	P91773 penaeus jap
5	63	65.6	1635	5 O17412	O17412 aedes aegypt
6	57	59.4	467	5 O15933	O15933 penaeus jap
7	53	55.2	688	5 Q8MS85	Q8MS85 drosophila
8	49	51.0	269	16 Q881H7	Q881H7 pseudomonas
9	47	49.0	467	5 Q8ITU3	Q8ITU3 penaeus van
10	46	47.9	484	5 Q9W032	Q9W032 drosophila
11	46	47.9	561	13 Q90711	Q90711 gallus gall
12	46	47.9	1016	2 Q8VRN4	Q8VRN4 rhodobacter
13	46	47.9	1219	13 Q90710	Q90710 gallus gall
14	45.5	47.4	258	2 Q9F8S0	Q9F8S0 escherichia
15	45.5	47.4	258	2 Q9F8R7	Q9F8R7 escherichia
16	45.5	47.4	258	16 Q8FEK7	Q8FEK7 escherichia

17	45	46.9	135	4 Q9NMJ5	Q9NMJ5 homo sapien
18	45	46.9	203	4 Q9B0B1	Q9B0B1 homo sapien
19	45	46.9	257	5 Q7YXW0	Q7YXW0 drosophila
20	45	46.9	329	16 Q89WQ5	Q89WQ5 bradyrhizob
21	45	46.9	368	5 Q9W2M5	Q9W2M5 drosophila
22	45	46.9	502	16 Q8BPF4	Q8BPF4 oceanobacti
23	44	45.8	275	2 Q8KX16	Q8KX16 uncultured
24	44	45.8	395	3 Q94467	Q94467 schizosach
25	44	45.8	405	16 Q812S7	Q812S7 bacillus ce
26	44	45.8	685	17 Q26481	Q26481 methanobact
27	44	45.8	636	17 Q9YF10	Q9YF10 aeropyrum p
28	44	45.8	958	16 Q8A070	Q8A070 bacteroides
29	44	45.8	1611	3 Q42854	Q42854 schizosach
30	43.5	45.3	438	16 Q7WNG0	Q7WNG0 rhodospirell
31	43	44.8	148	11 Q9DAH4	Q9DAH4 mus musculu
32	43	44.8	224	4 Q96M39	Q96M39 homo sapien
33	43	44.8	265	16 Q9X0W0	Q9X0W0 thermotoga
34	43	44.8	333	4 Q96SD4	Q96SD4 homo sapien
35	43	44.8	367	4 Q8N6Q2	Q8N6Q2 homo sapien
36	43	44.8	367	4 Q8IXT2	Q8IXT2 homo sapien
37	43	44.8	425	17 Q97W07	Q97W07 sulfobolus
38	43	44.8	442	5 Q96667	Q96667 drosophila
39	43	44.8	442	5 Q9W303	Q9W303 drosophila
40	43	44.8	452	2 Q8GFB6	Q8GFB6 rhodococcus
41	43	44.8	655	10 Q8H0K5	Q8H0K5 triticum ae
42	43	44.8	655	16 Q9RYM2	Q9RYM2 deinococcus
43	43	44.8	783	16 Q7WPE3	Q7WPE3 bordetella
44	43	44.8	783	16 Q7W3T4	Q7W3T4 bordetella
45	43	44.8	785	16 Q8XTS2	Q8XTS2 ralsstonia s

## ALIGNMENTS

RESULT 1  
ID Q9UER7 PRELIMINARY; PRT; 555 AA.  
AC Q9UER7;  
DT 01-MAY-2000 (TRENBLrel. 13. Created)  
DT 01-MAY-2000 (TRENBLrel. 13. Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25. Last annotation update)  
DE 98kDa HDM allergen.  
OS Dermatophagoides farinae (House-dust mite).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
OC Acariformes; Sarcoptiformes; Astigmata; Psoroptida; Analgoidee;  
OC Pyroglyphidae; Dermatophagoides.  
NCBI\_TaxID=6954;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Weber E.R., Hunter S., Stegman K., McGill C.;  
RT "Cloning and Characterization of a 98 kDa Allergen from  
RT Dermatophagoides farinae.";  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF178772; AAD52672.1; -;  
DR GO; GO:0005576; C:extracellular; IEA.  
DR GO; GO:0008061; F:chitin binding; IEA.  
DR GO; GO:0016798; F:hydrolase activity; acting on glycosyl bonds; IEA.  
DR GO; GO:0005975; F:carbohydrate metabolism; IEA.  
DR GO; GO:0006030; P:chitin metabolism; IEA.  
DR InterPro; IPR002557; Chitin bind. Pept.  
DR InterPro; IPR001223; Glyco\_hydro\_18.  
DR InterPro; IPR001579; Glyco\_hydro\_18A.  
DR Pfam; PF00704; Glyco\_hydro\_18; 1.  
DR ProDom; PD000471; Glyco\_hydro\_18; 1.  
DR SMART; SMO0494; ChitBD2; 1.  
DR SMART; SMO0636; Glyco\_18; 1.  
DR PROSITE; PS01095; CHITINASE\_18; 1.  
KW Glycosidase; Hydrolase.  
SQ SEQUENCE 555 AA; 63238 MW; 0E4564A1A459B30B CRC64;  
Query Match 100.0%; Score 96; DB 5; Length 555;  
Best Local Similarity 100.0%; Pred. No. 2.9e-07;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DAFPHGYLTAAVSPGK 18  
 DB 185 DAFPHGYLTAAVSPGK 202

RESULT 2  
 Q8MP05 PRELIMINARY; PRT; 2838 AA.  
 AC Q8MP05;  
 DT 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Chitinase precursor (EC 3.2.1.14).  
 GN CHIT5.  
 OS Tenebrio molitor (Yellow mealworm).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;  
 OC Tenebrionidae; Tenebrio.  
 OC NCBI\_TaxID=7067;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Royer V., Fraichard S., Bouhin H.;  
 RT "A Novel putative insect Chitinase with multiple catalytic domains :  
 RT hormonal regulation during metamorphosis."  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ487081; CAD31740.4; "-  
 DR GO: GO:0005576; C:extracellular; IEA.  
 DR GO: GO:0008061; F:chitin binding; IEA.  
 DR GO: GO:0008843; F:endochitinase activity; IEA.  
 DR GO: GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.  
 DR GO: GO:0005975; P:carbohydrate metabolism; IEA.  
 DR GO: GO:0006030; P:chitin metabolism; IEA.  
 DR InterPro: IPR002557; Chitin bind. Pera.  
 DR InterPro: IPR001223; Glyco\_hydro\_18.  
 DR InterPro: IPR001579; Glyco\_hydro\_18AS.  
 DR InterPro: IPR000634; S/T dehydratase\_BS.  
 DR Pfam: PF01607; CBM\_14; 5.  
 DR Pfam: PF00704; Glyco\_hydro\_18; 5.  
 DR ProDom: PD000471; Glyco\_hydro\_18; 5.  
 DR SMART: SM00494; ChitBD2; 5.  
 DR SMART: SM00636; Glyco\_18; 5.  
 DR PROSITE: PS01095; CHITINASE\_18; 3.  
 DR PROSITE: PS00165; DEHYDRATASE\_SER\_THR; 1.  
 DR Signal; Hydrolase; Glycosidase.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 21 2838 CHITINASE.  
 SQ SEQUENCE 2838 AA; 321407 MW; 608B3F2A8B9B937 CRC64;

Query Match 69.8%; Score 67; DB 5; Length 2838;  
 Best Local Similarity 76.5%; Pred. No. 0.1;  
 Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 AEPHGYLTAAVSPGK 18  
 DB 1645 AEPHGYLTAAVSPGK 1661

RESULT 3  
 Q8ISH5 PRELIMINARY; PRT; 431 AA.  
 AC Q8ISH5;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Chitinase.  
 OS Araneus ventricosus.  
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;  
 OC Araneomorphae; Entelegynae; Araneidae; Araneidae; Araneus.  
 OC NCBI\_TaxID=182803;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Han J.H., Kim S.R., Sohn H.D., Jin B.R.;  
 RT "Molecular cloning of a cDNA encoding the chitinase from the spider,  
 RT Araneus ventricosus."  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY120879; AAN39100.1; "-  
 DR GO: GO:0016787; F:hydrolase activity; IEA.  
 DR GO: GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro: IPR001223; Glyco\_hydro\_18.  
 DR InterPro: IPR001579; Glyco\_hydro\_18AS.  
 DR Pfam: PF00704; Glyco\_hydro\_18; 1.  
 DR ProDom: PD000471; Glyco\_hydro\_18; 1.  
 DR SMART: SM00636; Glyco\_18; 1.  
 DR PROSITE: PS01095; CHITINASE\_18; 1.  
 SQ SEQUENCE 431 AA; 47238 MW; 929439397B9BC923 CRC64;

Query Match 68.8%; Score 66; DB 5; Length 431;  
 Best Local Similarity 72.2%; Pred. No. 0.02;  
 Matches 13; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 DAFPHGYLTAAVSPGK 18  
 DB 178 DAFPHGYLTAAVSPGK 195

RESULT 4  
 P91773 PRELIMINARY; PRT; 527 AA.  
 AC P91773;  
 DT 01-MAY-1997 (TREMBlrel. 03, Created)  
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Fjchi-2.  
 OS Penaeus japonicus (Kuruma prawn).  
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
 OC Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;  
 OC Penaeidae; Marsupeneus.  
 OC NCBI\_TaxID=27405;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Matanabe T., Kono M.;  
 RT "Isolation of a cDNA Encoding a Chitinase Family Protein from  
 RT Cuticular Tissues of the Kuruma Prawn Penaeus japonicus."  
 RL Zool. 137: 1-6 (1996).  
 DR EMBL: D89751; BAAL4014.1; "-  
 DR GO: GO:0005576; C:extracellular; IEA.  
 DR GO: GO:0008061; F:chitin binding; IEA.  
 DR GO: GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.  
 DR GO: GO:0005975; P:carbohydrate metabolism; IEA.  
 DR GO: GO:0006030; P:chitin metabolism; IEA.  
 DR InterPro: IPR002557; Chitin bind. Pera.  
 DR InterPro: IPR001223; Glyco\_hydro\_18.  
 DR InterPro: IPR001579; Glyco\_hydro\_18AS.  
 DR InterPro: IPR001579; Glyco\_hydro\_18AS.  
 DR Pfam: PF01607; CBM\_14; 1.  
 DR Pfam: PF00704; Glyco\_hydro\_18; 1.  
 DR ProDom: PD000471; Glyco\_hydro\_18; 1.  
 DR SMART: SM00494; ChitBD2; 1.  
 DR SMART: SM00636; Glyco\_18; 1.  
 DR PROSITE: PS01095; CHITINASE\_18; 1.  
 KW Glycosidase; Hydrolase.  
 SQ SEQUENCE 527 AA; 59162 MW; B9CBAEAB8CDF8710 CRC64;

Query Match 65.6%; Score 63; DB 5; Length 527;  
 Best Local Similarity 80.0%; Pred. No. 0.077;  
 Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 AEPHGYLTAAVSP 16  
 DB 301 AEPHGYLTAAVSP 315

RESULT 5  
 O17412 PRELIMINARY; PRT; 1635 AA.  
 ID O17412



AC 017412;  
 DT 01-JAN-1998 (Tremblrel. 05, Created)  
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Probable chitinase 2 (EC 3.2.1.14).  
 GN CHIT2.  
 OS Aedes aegypti (yellowfever mosquito).  
 CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 CC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Aedes.  
 CC NCB1\_TaxID=7159;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98324849; PubMed=9662472;  
 RA de la Vega H., Specht C.A., Liu Y., Robbins P.W.;  
 RT "Chitinases are a multi-gene family in Aedes, Anopheles and  
 Drosophila.";  
 RL Insect Mol. Biol. 7:233-239(1998).  
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES OF N-  
 ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.  
 CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL  
 HYDROLASES).  
 CC EMBL; AF026492; AAB81850.1; -.  
 DR PIR; T14075; T14075.  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0008061; P:chitin binding; IEA.  
 DR GO; GO:0008843; F:endochitinase activity; IEA.  
 DR GO; GO:0016798; F:hydrolyase activity, acting on glycosyl bonds; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR GO; GO:0006032; P:chitin catabolism; IEA.  
 DR InterPro; IPR002557; Chitin bind PerA.  
 DR InterPro; IPR001223; Glyco\_hydro\_18.  
 DR InterPro; IPR001579; Glyco\_hydro\_18AS.  
 DR Pfam; PF01607; CBM\_14; 3.  
 DR ProDom; PD000471; Glyco\_hydro\_18; 3.  
 DR Pfam; PF00704; Glyco\_hydro\_18; 3.  
 DR SMART; SMO0636; Glyco\_18; 3.  
 DR PROSITE; PS01095; CHITINASE\_18; 3.  
 KW Hydrolyase; Glycosidase; Chitin degradation; Glycoprotein;  
 KW Multigene family.  
 FT CARBOHYD 132 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 322 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 463 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 749 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 890 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1338 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1479 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1635 AA; 185993 MW; EHL16F83AAC129FA CRC64;

Query Match 65.6%; Score 63; DB 5; Length 1635;  
 Best Local Similarity 70.6%; Pred. No. 0.26;  
 Matches 12; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 AFPHGYLITAAVSPK 18  
 Db 393 AFPHGYLITAAVSPK 409  
 RESULT 6  
 OLS993 PRELIMINARY; PRT; 467 AA.  
 AC 015993;  
 DT 01-JAN-1998 (Tremblrel. 05, Created)  
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Pjchl-3.  
 OS Penaeus japonicus (Kuruma prawn).  
 CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Malacostraca;  
 CC Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;  
 CC Penaeidae; Mazuspinaeus.  
 CC NCB1\_TaxID=7405;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Matanabe T., Kono M., Aida K., Nagasawa H.;  
 RT "Purification and molecular cloning of a chitinase expressed in the  
 RT hepatopancreas of the penaeid prawn Penaeus japonicus.";  
 RL Biochim. Biophys. Acta 0:0-0(1997).  
 DR EMBL; AB008027; BAA22854.1; -.  
 DR HSP; P07254; ICTN.  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0008061; P:chitin binding; IEA.  
 DR GO; GO:0016787; F:hydrolyase activity; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR GO; GO:0006030; P:chitin metabolism; IEA.  
 DR InterPro; IPR002557; Chitin bind PerA.  
 DR InterPro; IPR001223; Glyco\_hydro\_18.  
 DR Pfam; PF01607; CBM\_14; 1.  
 DR Pfam; PF00704; Glyco\_hydro\_18; 1.  
 DR ProDom; PD000471; Glyco\_hydro\_18; 1.  
 DR SMART; SMO0494; ChitBD2; 1.  
 DR SMART; SMO0636; Glyco\_18; 1.  
 DR PROSITE; PS01095; CHITINASE\_18; 1.  
 SQ SEQUENCE 467 AA; 51765 MW; 499F7095774CA445 CRC64;

Query Match 59.4%; Score 57; DB 5; Length 467;  
 Best Local Similarity 70.6%; Pred. No. 0.67;  
 Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 AFPHGYLITAAVSPK 18  
 Db 150 ALQPEGLTAAVSPK 166

RESULT 7  
 Q8MS85 PRELIMINARY; PRT; 688 AA.  
 ID Q8MS85;  
 AC Q8MS85;  
 DT 01-OCT-2002 (Tremblrel. 22, Created)  
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE LP04696P.  
 GN BCDNA:LP04696.  
 OS Drosophila melanogaster (fruit fly).  
 CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 CC NCB1\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkely;  
 RA Stapleton W., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
 RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,  
 RA Miranda A., Mungall C.J., Nuno J., Paclab J., Paredas V., Park S.,  
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
 RA Cejner S.;  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY119021; AAMS0881.1; -.  
 DR FLYBase; FBgn063117; BCDNA:LP04696.  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0008061; F:chitin binding; IEA.  
 DR GO; GO:0016787; F:hydrolyase activity; IEA.  
 DR GO; GO:0006030; P:chitin metabolism; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR InterPro; IPR002557; Chitin bind PerA.  
 DR InterPro; IPR001223; Glyco\_hydro\_18.  
 DR Pfam; PF01607; CBM\_14; 1.  
 DR Pfam; PF00704; Glyco\_hydro\_18; 2.  
 DR ProDom; PD000471; Glyco\_hydro\_18; 2.  
 DR SMART; SMO0494; ChitBD2; 1.  
 DR SMART; SMO0636; Glyco\_18; 1.  
 SQ SEQUENCE 688 AA; 78069 MW; 25845A861098492D CRC64;

Query Match 55.2%; Score 53; DB 5; Length 688;  
 Best Local Similarity 55.6%; Pred. No. 4.6;

Matches 10; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 1 DAFEPHYLLTAAVSPGK 18  
Db 461 EAFKENGILTAAVSPGK 478

## RESULT 8

Q881H7 PRELIMINARY; PRT; 269 AA.  
AC Q881H7;  
DT 01-JUN-2003 (TREMBlrel. 24, Created)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Transcriptional regulator, Arac family.  
GN PF0302.  
OS Pseudomonas putida (strain KT2440).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=160488;  
RN [1]  
RE SEQUENCE FROM N.A.  
RX MEDLINE=2423060; PubMed=12534463;  
RA Nelson K.E., Weinl C., Paulsen I.T., Dodson R.U., Hilbert H.,  
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,  
RA Brinkac L., Bearan M., DeBoy R.T., Daugherty S., Kolonay J.,  
RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,  
RA Chris Lee P., Holtzapfel B., Scanlan D., Tran K., Moazzaz A.,  
RA Utechtack T., Rizzo W., Lee K., Kosack D., Moesti D., Medler H.,  
RA Lauber U., Stjepandic D., Honeisel U., Straetz M., Heim S.,  
RA Kiewitz C., Eisen J., Timmis K.N., Duesterhoeft A., Tuenmler B.,  
RA Fraser C.M.;  
RT "Complete genome sequence and comparative analysis of the  
RT metabolically versatile Pseudomonas putida KT2440.";  
RL Envision. Microbiol. 4:799-808(2002).  
DR EMBL; AE016785; AAN68630.1; -.  
DR TIGR; PF03022; -.  
DR GO; GO:0005622; C:intracellular; IEA.  
DR GO; GO:0003700; F:transcription factor activity; IEA.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR InterPro; IPR003313; Arac\_binding.  
DR InterPro; IPR000005; HTMArc.  
DR Pfam; PF02311; Arac\_binding; 1.  
DR Pfam; PF00165; HTH\_Arac; 2.  
DR PRINTS; PR00032; HTMArc.  
DR PROSITE; PS00041; HTH\_ARAC\_FAMILY\_1; 1.  
DR PROSITE; PS01124; HTH\_ARAC\_FAMILY\_2; 1.  
KW complete proteome.  
SQ SEQUENCE 269 AA; 29189 MW; C05C6F2AC8C2C03 CRC64;

Query Match 51.0%; Score 49; DB 16; Length 269;  
Best Local Similarity 56.2%; Pred. No. 7.8;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 3 FEPHYLLTAAVSPGK 18  
Db 212 FTFHYLLTAAVSPGK 227

## RESULT 9

Q81TJ3 PRELIMINARY; PRT; 467 AA.  
AC Q81TJ3;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Chitinase.  
OS Penaeus vannamei (Penaeid shrimp) (European white shrimp).  
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
OC Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;  
OC Penaeidae; Litopenaeus.  
OX NCBI\_TaxID=66699;  
RN [1]

RP SEQUENCE FROM N.A.

RA Chang R.C.;  
RT "Cloning and characterization of a cDNA encoding a chitinase from  
RT hepatopancreas of the Penaeus vannamei (Crustacea, Decapoda).";  
RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AF315689; AAN74647.1; -.  
DR GO; GO:0005576; C:extracellular; IEA.  
DR GO; GO:0008061; F:chitin binding; IEA.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR GO; GO:0006030; P:chitin metabolism; IEA.  
DR InterPro; IPR002557; Chitin\_bind\_Pera.  
DR InterPro; IPR001223; Glyco\_hydro\_18.  
DR InterPro; IPR001579; Glyco\_hydro\_18AS.  
DR Pfam; PF01607; CSM\_14; 1.  
DR Pfam; PF00704; Glyco\_hydro\_18; 1.  
DR ProDom; PD000471; Glyco\_hydro\_18; 1.  
DR SMART; SM00494; ChEBD2; 1.  
DR SMART; SM00636; Glyco\_18; 1.  
DR PROSITE; PS01095; CHITINASE\_18; 1.  
SQ SEQUENCE 467 AA; 51958 MW; 734A830C6F47F4CD CRC64;

Query Match 49.0%; Score 47; DB 5; Length 467;  
Best Local Similarity 64.7%; Pred. No. 30;  
Matches 11; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 AEPHYLLTAAVSPGK 18  
Db 150 ALHAEGLTAAVSPGK 166

## RESULT 10

Q9W092 PRELIMINARY; PRT; 484 AA.  
AC Q9W092; O17421;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Probable chitinase 2 precursor (EC 3.2.1.14) (UD28264p).  
GN CHT2 OR CG2054.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RE SEQUENCE FROM N.A.  
RX STRAIN=Berkeley;  
RC MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blake J.R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter J.G., Helt G., Nelson C.R., McKis G.L.G.,  
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu I., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhattacharya D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Flosser C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck C.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kesterson J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,



DR PROSITE; PS50046; PHYTOCHROME\_2; 1.  
 SQ PHYTOCHROME.  
 KM SEQUENCE 1016 AA; 110051 MW; F86F8057F0F18DB3 CRC64;  
 RT Query Match 47.9%; Score 46; DB 2; Length 1016;  
 Best Local Similarity 47.1%; Pred. No. 1e+02;  
 Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
 QY 2 AFEPHGYLTAAVSPG 18  
 DB 108 AIOPHGALMTARADSGR 124  
 RESULT 13  
 ID Q90710 PRELIMINARY; PRT; 1219 AA.  
 AC Q90710;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE AE2-1 anion exchanger.  
 DE AE2.  
 OS Gallus gallus (Chicken).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Proventriculus;  
 RX MEDLINE=96224107; PubMed=8621532;  
 RA Cox K.H., Adair-Kirk T.V., Cox J.V.;  
 RT "Variant AE2 anion exchanger transcripts accumulate in multiple cell  
 RT types in the chicken gastric epithelium.";  
 RL J Biol. Chem. 271:8995-8902(1996).  
 DR EMBL; U48889; AAC59881.1; -.  
 DR HSP; P02730; IRT0.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005452; F:inorganic anion exchanger activity; IEA.  
 DR GO; GO:0006820; P:anion transport; IEA.  
 DR InterPro; IPR001717; Anion exchange.  
 DR InterPro; IPR003020; HCO3\_cotranspct.  
 DR Pfam; PF00955; HCO3\_cotranspct. 1.  
 DR PRINTS; PRO1231; HCO3TRANSPORT.  
 DR TIGRFAMs; TIGR00834; ae; 1.  
 DR PROSITE; PS00219; ANION\_EXCHANGER\_1; 1.  
 DR PROSITE; PS00220; ANION\_EXCHANGER\_2; 1.  
 DR PROSITE; 1219 AA; 135288 MW; 25F42A73C3483B21 CRC64;  
 SQ SEQUENCE  
 Query Match 47.9%; Score 46; DB 13; Length 1219;  
 Best Local Similarity 64.3%; Pred. No. 1.2e+02;  
 Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 4 EEPHYLTAAVSPG 17  
 DB 181 EPHGATAPAAASPG 194  
 RESULT 14  
 ID Q9F880 PRELIMINARY; PRT; 258 AA.  
 AC Q9F880;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 DE YGBM.  
 OS Escherichia coli.  
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=DEC 12e;  
 RX MEDLINE=20444184; PubMed=10986240;  
 RA Herbelin C.J., Chirillo S.C., Melnick K.A., Whitlam T.S.;  
 RT "Gene Conservation and Loss in the mutS-tpos Genomic Region of  
 RT Pathogenic Escherichia coli.";  
 RL J. Bacteriol. 182:5381-5390(2000).  
 DR EMBL; AF24208; AAG14970.1; -.  
 DR GO; GO:0005622; C:intracellular; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR GO; GO:0004519; F:endonuclease activity; IEA.  
 DR GO; GO:0006281; F:DNA repair; IEA.  
 DR InterPro; IPR001719; AP\_endonuclease2.  
 DR Pfam; PF01261; AP\_endonuc\_2; 1.  
 DR Hypothetical protein.  
 SQ SEQUENCE 258 AA; 29218 MW; EFF7A3C904393AED CRC64;  
 Query Match 47.4%; Score 45.5; DB 2; Length 258;  
 Best Local Similarity 55.6%; Pred. No. 28;  
 Matches 10; Conservative 2; Mismatches 5; Indels 1; Gaps 1;  
 QY 1 DAFEPHG-YLTAAVSPG 17  
 DB 131 DRFAPHGKILVIALSPG 148

RESULT 15  
 ID Q9F87 PRELIMINARY; PRT; 258 AA.  
 AC Q9F87;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 DE YGBM.  
 OS Escherichia coli.  
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DEC 1a;  
 RX MEDLINE=20444184; PubMed=10986240;  
 RA Herbelin C.J., Chirillo S.C., Melnick K.A., Whitlam T.S.;  
 RT "Gene Conservation and Loss in the mutS-tpos Genomic Region of  
 RT Pathogenic Escherichia coli.";  
 RL J. Bacteriol. 182:5381-5390(2000).  
 DR EMBL; AF24209; AAG14976.1; -.  
 DR GO; GO:0005622; C:intracellular; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR GO; GO:0004519; F:endonuclease activity; IEA.  
 DR GO; GO:0006281; F:DNA repair; IEA.  
 DR InterPro; IPR001719; AP\_endonuclease2.  
 DR Pfam; PF01261; AP\_endonuc\_2; 1.  
 DR Hypothetical protein.  
 SQ SEQUENCE 258 AA; 29146 MW; 76B57850B070273C CRC64;  
 Query Match 47.4%; Score 45.5; DB 2; Length 258;  
 Best Local Similarity 55.6%; Pred. No. 28;  
 Matches 10; Conservative 2; Mismatches 5; Indels 1; Gaps 1;  
 QY 1 DAFEPHG-YLTAAVSPG 17  
 DB 131 DRFAPHGKILVIALSPG 148

Search completed: March 22, 2004, 06:59:18  
 Job time : 5.62201 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:30:23 ; Search time 5.47304 Seconds  
(without alignments)  
929.256 Million cell updates/sec

Title: US-09-662-293-6

Perfect score: 96

Sequence: 1 DAEPHRYLITRAVSPGK 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1980s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	96	100.0	18	AA52515	AAY52515 House dus
2	96	100.0	18	AAU96319	Aau96319 Der HMW-m
3	96	100.0	536	AA52525	Aay52525 House dus
4	96	100.0	536	AAU96329	Aau96329 Der HMW-m
5	96	100.0	555	AA52522	Aay52522 House dus
6	96	100.0	555	AAU96327	Aau96327 Der HMW-m
7	96	100.0	555	AAU96328	Aau96328 Der HMW-m
8	96	86.5	490	AA52535	Aay52535 D. pteron
9	83	86.5	490	AAU96339	Aau96339 Der HMW-m
10	83	86.5	509	AA52533	Aay52533 D. pteron
11	83	86.5	509	AAU96337	Aau96337 Der HMW-m
12	83	86.5	509	AAU96338	Aau96338 Der HMW-m
13	83	65.6	527	ABP72624	Abp72624 Prwm chl
14	69	61.5	508	ABBE6399	Abbe6399 Drosophi
15	57	59.4	467	ABP72634	Abp72634 Prwm chl
16	53	55.2	305	ABBE6689	Abbe6689 Drosophi
17	52	54.2	458	ABBE6690	Abbe6690 Drosophi
18	46	47.9	484	ABBS6787	Abbs6787 Drosophi
19	45.5	47.4	389	AAW85454	Aaw85454 An intein
20	45.5	47.4	389	AAW87550	Aaw87550 Intein (p
21	45.5	47.4	1523	AAW87554	Aaw87554 A Tfu DNA
22	45	46.9	135	AAW87713	Aaw87713 Human pro
23	45	46.9	140	AAW28577	Aaw28577 Secreted
24	45	46.9	140	AAU39049	Aau39049 Human sec
25	45	46.9	140	ABBS5758	Abb5758 Human pol

26	45	46.9	165	7	AA030424	AA030424 Human sec
27	45	46.9	203	5	ABU10918	ABU10918 Human sec
28	45	46.9	203	6	ABG99928	ABG99928 Human nov
29	45	46.9	204	7	ADD45619	Add45619 Human pro
30	45	46.9	219	5	ABJ10925	Abj10925 Human sec
31	45	46.9	368	4	ABBE5118	Abbe5118 Drosophi
32	44	45.8	103	3	AAW41243	Aaw41243 Human ORF
33	44	45.8	103	5	ABP06845	Abp06845 Human ORF
34	43.5	45.3	391	4	ABG99091	ABG99091 Thermoco
35	43	44.8	114	4	AAW49436	Aaw49436 Human pro
36	43	44.8	224	6	ADA55557	Ada55557 Human pro
37	43	44.8	442	3	AAW07177	Aaw07177 Drosophi
38	43	44.8	442	4	ABBS8384	Abbs8384 Drosophi
39	42.5	44.3	687	7	ADB37628	Adb37628 Neutral ch
40	42	43.8	113	4	AAW43232	Aaw43232 Propionib
41	42	43.8	113	6	ABM39751	Abm39751 Human NOV
42	42	43.8	302	5	ABU65207	Abu65207 Human NOV
43	42	43.8	500	5	ABU10465	Abu10465 Ciera reco
44	42	43.8	707	6	ABP97753	Abp97753 Amino aci
45	41	42.7	29	4	ABG00774	Abg00774 Novel hum

## ALIGNMENTS

RESULT 1  
AA52515  
ID AAY52515 standard; peptide; 18 AA.  
XX  
XX AAY52515;  
AC  
XX  
DT 22-FEB-2000 (first entry)  
XX  
XX House dust mite allergen protein (map) A/B fragment map(5).  
DE  
XX  
XX Mite allergen protein; map; high molecular weight; HMW-map; allergy;  
KW house dust mite; IGE; immunoglobulin E; allergen; map; map;  
KW hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
KW canine; veterinary; antibody; vaccine; immunisation.  
XX  
XX  
OS Dermatophagoides farinae.  
XX  
XX  
PN W09954349-A2.  
XX  
PF 16-APR-1999; 99WO-US008524.  
XX  
PR 17-APR-1998; 98US-00062013.  
PR 13-MAY-1998; 98US-0085295P.  
PR 02-SEP-1998; 98US-0098909P.  
XX  
PA (HESK-) HESKA CORP.  
XX  
PI McCalli CA, Hunter SW, Weber ER;  
XX  
XX WPI; 2000-052700/04.  
DR  
XX  
PT Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
PT used to modify an animals' hypersensitivity to mite allergens.  
XX  
PS Claim 3; Page 69; 154pp; English.  
XX  
XX Sequences AAY52510-Y52522 represent proteolytic fragments of  
CC Dermatophagoides farinae high molecular weight mite allergen protein (HMW  
CC map) composition. The HMW-map composition was isolated from a D. farinae  
CC homogenate by gel filtration, with each fraction being analysed for the  
CC presence of proteins that bound to IGE present in mite-allergic dog  
CC antisera. The HMW-map composition comprises mapA (a 109 kD protein) and  
CC mapB (98 kD). Mite allergenic proteins and peptides, and nucleic acids  
CC encoding them, may be used in therapeutic compositions to modify an  
CC animal's hypersensitivity reaction to mite allergens. Animals that may be  
CC treated include mammals and birds, especially felines, canines, equines,

CC humans, other pets, and work or domestic animals. The proteins or  
 CC fragments may also be used to diagnose allergies via a skin test. The  
 CC proteins and peptides can also be used to raise antibodies, which have a  
 CC variety of potential uses. For example, they can be used as vaccines to  
 CC passively immunise animals against dust mite hypersensitivity, as  
 CC positive controls in test kits and as tools to recover desired dust mite  
 CC allergens from a mixture of proteins

XX Sequence 18 AA;

Query Match 100.0%; Score 96; DB 3; Length 18;

Best Local Similarity 100.0%; Pred. No. 1e-09;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DAFEPHGYLITAAVSPGK 18  
 Db 1 DAFEPHGYLITAAVSPGK 18

RESULT 2

AAU96319 standard; peptide; 18 AA.

XX AAU96319;

XX 15-JUL-2002 (first entry)

XX Der HMW-map polypeptide #6.

XX Der HMW-map; American house dust mite; anti-allergic; mite; IGE;

XX mite allergenic protein; immunoglobulin E; hypersensitivity;

XX immunocomplex formation.

XX Dermatophagoides farinae.

XX WO200222807-A2.

XX 21-MAR-2002.

XX 14-SEP-2001; 2001WO-US028730.

XX 14-SEP-2000; 2000US-00662293.

XX (HESK-) HESKA CORP.

XX Mccall CA, Hunter SW, Weber ER;

XX WPI; 2002-351888/38.

XX New mite allergenic protein isolated from Dermatophagoides, designated

XX Der HMW-map protein, useful as a vaccine for treating mite allergy.

XX Claim 12; Page 70; 161pp; English.

XX The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic  
 CC acid. The Der HMW-map protein is useful for eliciting an immune response  
 CC against Der HMW-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitising a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig) E or Der HMW-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HMW-map polypeptides of the invention

XX Sequence 18 AA;

Query Match 100.0%; Score 96; DB 5; Length 18;

Best Local Similarity 100.0%; Pred. No. 1e-09;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DAFEPHGYLITAAVSPGK 18  
 Db 1 DAFEPHGYLITAAVSPGK 18

RESULT 3

AAU52525 standard; protein; 536 AA.

XX AAU52525;

XX 22-FEB-2000 (first entry)

XX House dust mite (D. farinae) mite allergen protein (map) Pderf98-536.

XX Mite allergen protein; map; high molecular weight; HMW-map; allergy;

XX house dust mite; IGE; immunoglobulin E; allergen; map; mapB;

XX hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;

XX canine; veterinary; antibody; vaccine; immunisation.

XX Dermatophagoides farinae.

XX WO9954349-A2.

XX 16-APR-1999; 99NO-US008524.

XX 17-APR-1998; 98US-00062013.

XX 13-MAY-1998; 98US-0085295P.

XX 02-SEP-1998; 98US-0098909P.

XX (HESK-) HESKA CORP.

XX Mccall CA, Hunter SW, Weber ER;

XX WPI; 2000-052700/04.

XX N-PSDB; AA238579, AA238580.

XX Novel high molecular weight Dermatophagoides nucleic acid polypeptides

XX used to modify an animal's hypersensitivity to mite allergens.

XX Claim 3; Page 125-127; 154pp; English.

XX This sequence represents Dermatophagoides farinae mite allergen protein  
 CC (map) Pderf98-536, the mature form of Pderf98-555 (AAU52523). Pderf98-536  
 CC has a molecular weight of 98 kD, comprising 536 amino acids, and is a  
 CC component of the Dermatophagoides farinae high molecular weight mite  
 CC allergen protein (HMW-map) composition. The HMW-map composition was  
 CC isolated from a D. farinae homogenate by gel filtration, with each  
 CC fraction being analysed for the presence of proteins that bound to IgE  
 CC present in mite-allergic dog antisera. Mite allergenic proteins and  
 CC peptides, and nucleic acids encoding them, may be used in therapeutic  
 CC compositions to modify an animal's hypersensitivity reaction to mite  
 CC allergens. Animals that may be treated include mammals and birds,  
 CC especially felines, canines, equines, humans, other pets, and work or  
 CC domestic animals. The proteins or fragments may also be used to diagnose  
 CC allergies via a skin test. The proteins and peptides can also be used to  
 CC raise antibodies, which have a variety of potential uses. For example,  
 CC they can be used as vaccines to passively immunise animals against dust  
 CC mite hypersensitivity, as positive controls in test kits and as tools to  
 CC recover desired dust mite allergens from a mixture of proteins

XX Sequence 536 AA;

Query Match 100.0%; Score 96; DB 3; Length 536;

Best Local Similarity 100.0%; Pred. No. 6.7e-08;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 DAFEPHGYLITAAVSPGK 18

Db 166 DAFEPHGYLTLTAAVSPGX 183

RESULT 4  
AAU96329 standard; protein; 536 AA.

AAU96329;

15-JUL-2002 (first entry)

Der HMW-map polypeptide #16.

Der HMW-map; American house dust mite; antiallergic; mite; IgE;  
mite allergenic protein; immunoglobulin E; hypersensitivity;  
immunocomplex formation.

Dermatophagoides farinae.

WO200222807-A2.

21-MAR-2002.

14-SEP-2001; 2001MO-US028730.

14-SEP-2000; 2000US-00662293.

(HESK-) HESKA CORP.

Mccall CA, Hunter SW, Weber ER;

WPI; 2002-351888/38.

N-PSDB; ABK69575.

New mite allergenic protein isolated from Dermatophagoides, designated

Der HMW-map protein, useful as a vaccine for treating mite allergy.

Claim 12; Page 125-127; 161pp; English.

The invention relates to an isolated mite allergenic protein of  
Dermatophagoides, designated Der HMW-map protein, and its related nucleic  
acid. The Der HMW-map protein is useful for eliciting an immune response  
against Der HMW-map protein. The protein or a reagent comprising a non-  
proteinaceous epitope is useful for identifying an animal (e.g., dog,  
cat) susceptible to or having an allergic response to a mite. A  
therapeutic composition is useful for desensitizing a host animal to an  
allergic response to a mite. The DNA and protein can be used in the  
detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
of immunoglobulin (Ig)E or Der HMW-map protein activity associated with a  
disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
binding of proteins to IgE, to prevent immunocomplex formation, thus  
reducing hypersensitivity responses to mite allergens, and as vaccines  
against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
represent Der HMW-map polypeptides of the invention

Sequence 536 AA;

Query Match 100.0%; Score 96; DB 5; Length 536;  
Best Local Similarity 100.0%; Pred. No. 6,7e-08;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DAFEPHGYLTLTAAVSPGX 18  
166 DAFEPHGYLTLTAAVSPGX 183

RESULT 5  
AAU96323 standard; protein; 555 AA.  
AAU96323;  
AAU96323;

DT 22-FEB-2000 (first entry)

House dust mite (D. farinae) mite allergen protein (map) Pderf98-555.

Mite allergen protein; map; high molecular weight; HMW-map; allergy;

house dust mite; IgE; immunoglobulin E; allergen; mapA; mapB;

hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;

canine; veterinary; antibody; vaccine; immunisation.

Dermatophagoides farinae.

Key Location/Qualifiers

FT Peptide 1..19 /note="Signal peptide"

FT Protein 20..555 /note="Mature Pderf98-555"

WO9954349-A2.

28-OCT-1999.

16-APR-1999; 99WO-US008524.

17-APR-1998; 98US-00062013.

13-MAY-1998; 98US-0085295P.

02-SEP-1998; 98US-0098909P.

(HESK-) HESKA CORP.

Mccall CA, Hunter SW, Weber ER;

WPI; 2000-052700/04.

N-PSDB; AAZ38575, AAZ38576, AAZ38577, AAZ38578.

Novel high molecular weight Dermatophagoides nucleic acid polypeptides

used to modify an animals' hypersensitivity to mite allergens.

Claim 3; Page 111-113; 154pp; English.

This sequence represents Dermatophagoides farinae mite allergen protein  
(map) Pderf98-555. Pderf98-555 has a molecular weight of 98 kD,  
comprising 555 amino acids, and is a component of the Dermatophagoides  
farinae high molecular weight mite allergen protein (HMW-map)  
composition. The HMW-map composition was isolated from a D. farinae  
homogenate by gel filtration, with each fraction being analysed for the  
presence of proteins that bound to IgE present in mite-allergic dog  
antisera. Mite allergenic proteins and peptides, and nucleic acids  
encoding them, may be used in therapeutic compositions to modify an  
animal's hypersensitivity reaction to mite allergens. Animals that may be  
treated include mammals and birds, especially felines, canines, equines,  
humans, other pets, and work or domestic animals. The proteins or  
fragments may also be used to diagnose allergies via a skin test. The  
proteins and peptides can also be used to raise antibodies, which have a  
variety of potential uses. For example, they can be used as vaccines to  
passively immunise animals against dust mite hypersensitivity, as  
positive controls in test kits and as tools to recover desired dust mite  
allergens from a mixture of proteins

Sequence 555 AA;

Query Match 100.0%; Score 96; DB 3; Length 555;  
Best Local Similarity 100.0%; Pred. No. 7e-08;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DAFEPHGYLTLTAAVSPGX 18  
185 DAFEPHGYLTLTAAVSPGX 202

RESULT 6  
AAU96327 standard; protein; 555 AA.  
AAU96327  
AAU96327

AC AAU96327;  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX  
 DE Der HMW-map polypeptide #14.  
 XX  
 KW Der HMW-map; American house dust mite; anti-allergic; mite; IGE;  
 XX mite allergenic protein; immunoglobulin E; hypersensitivity;  
 XX immunocomplex formation.  
 XX  
 OS Dermatophagoides farinae.  
 XX  
 PN MO200222807-A2.  
 XX  
 PD 21-MAR-2002.  
 XX  
 PE 14-SEP-2001; 2001WO-US028730.  
 XX  
 PR 14-SEP-2000; 2000US-00662293.  
 XX  
 PA (HESK-) HESKA CORP.  
 XX  
 PI Mccall CA, Hunter SW, Weber ER;  
 XX  
 DR WPI; 2002-351888/38.  
 DR N-PSDB; ABK69571.  
 XX  
 PT New mite allergenic protein isolated from Dermatophagoides, designated  
 XX Der HMW-map protein, useful as a vaccine for treating mite allergy.  
 XX  
 PS Claim 12; Page 114-116; 161pp; English.  
 XX  
 CC The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic  
 CC acid. The Der HMW-map protein is useful for eliciting an immune response  
 CC against Der HMW-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitizing a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig) E or Der HMW-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
 CC binding of proteins to IGE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HMW-map polypeptides of the invention  
 XX  
 SQ Sequence 555 AA;

Query Match 100.0%; Score 96; DB 5; Length 555;  
 Best Local Similarity 100.0%; Pred. No. 7e-08;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DAFEPHGYLTLAAVSPGK 18  
 DB 185 DAFEPHGYLTLAAVSPGK 202

RESULT 7  
 AAU96328  
 ID AAU96328 standard; protein; 555 AA.  
 XX  
 AC AAU96328;  
 XX  
 DE 15-JUL-2002 (first entry)  
 XX  
 DE Der HMW-map polypeptide #15.  
 XX  
 KW Der HMW-map; American house dust mite; anti-allergic; mite; IGE;  
 XX mite allergenic protein; immunoglobulin E; hypersensitivity;  
 XX immunocomplex formation.

OS Dermatophagoides farinae.  
 XX  
 PN MO200222807-A2.  
 XX  
 PD 21-MAR-2002.  
 XX  
 PE 14-SEP-2001; 2001WO-US028730.  
 XX  
 PR 14-SEP-2000; 2000US-00662293.  
 XX  
 PA (HESK-) HESKA CORP.  
 XX  
 PI Mccall CA, Hunter SW, Weber ER;  
 XX  
 DR WPI; 2002-351888/38.  
 DR N-PSDB; ABK69573.  
 XX  
 PT New mite allergenic protein isolated from Dermatophagoides, designated  
 XX Der HMW-map protein, useful as a vaccine for treating mite allergy.  
 XX  
 PS Claim 12; Page 120-122; 161pp; English.  
 XX  
 CC The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic  
 CC acid. The Der HMW-map protein is useful for eliciting an immune response  
 CC against Der HMW-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitizing a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig) E or Der HMW-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
 CC binding of proteins to IGE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HMW-map polypeptides of the invention  
 XX  
 SQ Sequence 555 AA;

Query Match 100.0%; Score 96; DB 5; Length 555;  
 Best Local Similarity 100.0%; Pred. No. 7e-08;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DAFEPHGYLTLAAVSPGK 18  
 DB 185 DAFEPHGYLTLAAVSPGK 202

RESULT 8  
 AAU52535  
 ID AAU52535 standard; protein; 490 AA.  
 XX  
 AC AAU52535;  
 XX  
 DT 06-AUG-2003 (revised)  
 DT 22-FEB-2000 (first entry)  
 XX  
 DE D. pteronyssinus 98 kD mite allergen protein (map) PDerp98-490.  
 XX  
 KW Mite allergen protein; map; high molecular weight; HMW-map; allergy;  
 XX house dust mite; IGE; immunoglobulin E; allergen; mapB;  
 KW hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
 XX canine; veterinary; antibody; vaccine; immunisation.  
 XX  
 OS Dermatophagoides pteronyssinus.  
 XX  
 PN Key Location/Qualifiers  
 FT Modified-site 115..117  
 FT Modified-site /note="Asn is N-glycosylated"  
 FT Modified-site 240..242  
 FT Modified-site /note="Asn is N-glycosylated"  
 XX



PN W09954349-A2.  
 XX  
 XX 28-OCT-1999.  
 PD  
 XX  
 PF 16-APR-1999; 99WO-US008524.  
 XX  
 XX 17-APR-1998; 98US-00062013.  
 PR 13-MAY-1998; 98US-0085295P.  
 PR 02-SEP-1998; 98US-0098909P.  
 XX  
 PA (HESK-) HESKA CORP.  
 PI Mccall CA, Hunter SW, Weber ER;  
 XX  
 XX WPI: 2000-052700/04.  
 DR N-PSDB; AA238589; AA238590.  
 XX  
 XX Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
 PT used to modify an animals' hypersensitivity to mite allergens.  
 XX  
 XX Claim 3; Page 147-149; 154pp; English.  
 XX  
 XX This sequence represents Dermatophagoides pteronyssinus mite allergen  
 CC protein (map) Pderp98-490, the mature form of pderp98-509. Pderp98-490  
 CC has a molecular weight of 98 kD, comprising 490 amino acids, and has a  
 CC high degree of homology with the D. Farinae mature 98 kD allergen, mapB  
 CC (AAV5525). Nucleic acid molecules encoding Pderp98-490 were isolated  
 CC from a D. pteronyssinus cDNA library by hybridisation with a probe  
 CC encoding the D. farinae high molecular weight map (HWM-map) composition.  
 CC Mite allergenic proteins and peptides, and nucleic acids encoding them,  
 CC may be used in therapeutic compositions to modify an animal's  
 CC hypersensitivity reaction to mite allergens. Animals that may be treated  
 CC include mammals and birds, especially felines, canines, equines, humans,  
 CC other pets, and work or domestic animals. The proteins or fragments may  
 CC also be used to diagnose allergies via a skin test. The proteins and  
 CC peptides can also be used to raise antibodies, which have a variety of  
 CC potential uses. For example, they can be used as vaccines to passively  
 CC immunise animals against dust mite hypersensitivity, as positive controls  
 CC in test kits and as tools to recover desired dust mite allergens from a  
 CC mixture of proteins. (Updated on 06-AUG-2003 to correct OS field.)  
 CC  
 XX  
 SQ Sequence 490 AA;  
 Query Match 86.5%; Score 83; DB 3; Length 490;  
 Best Local Similarity 88.9%; Pred. No. 1.1e-05;  
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 DAFEPHGILTTAAVSPGK 18  
 :|||||  
 Db 166 EAFEPFGILTTAAVSPGK 183  
 RESULT 9  
 AAU96339  
 ID AAU96339 standard; protein; 490 AA.  
 XX  
 AC AAU96339;  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX  
 DE Der HWM-map polypeptide #26.  
 XX  
 XX Der HWM-map; American house dust mite; anti-allergic; mite; IGE;  
 KM mite allergenic protein; immunoglobulin E; hypersensitivity;  
 KM immunocomplex formation.  
 XX  
 OS Dermatophagoides farinae.  
 XX  
 PN W0200222807-A2.  
 PD  
 XX 21-MAR-2002.  
 PF  
 XX 14-SEP-2001; 2001WO-US028730.

XX  
 PR 14-SEP-2000; 2000US-00662293.  
 XX  
 XX (HESK-) HESKA CORP.  
 PA Mccall CA, Hunter SW, Weber ER;  
 XX  
 XX WPI: 2002-351888/38.  
 DR N-PSDB; ABK69585.  
 XX  
 PT New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HWM-map protein, useful as a vaccine for treating mite allergy.  
 XX  
 XX Claim 12; Page 144-146; 161pp; English.  
 XX  
 XX The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HWM-map protein, and its related nucleic  
 CC acid. The Der HWM-map protein is useful for eliciting an immune response  
 CC against Der HWM-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitising a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HWM-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig) E or Der HWM-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HWM-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HWM-map polypeptides of the invention  
 XX  
 XX  
 SQ Sequence 490 AA;  
 Query Match 86.5%; Score 83; DB 5; Length 490;  
 Best Local Similarity 88.9%; Pred. No. 1.1e-05;  
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 DAFEPHGILTTAAVSPGK 18  
 :|||||  
 Db 166 EAFEPFGILTTAAVSPGK 183  
 RESULT 10  
 AAU52533  
 ID AAU52533 standard; protein; 509 AA.  
 XX  
 AC AAU52533;  
 XX  
 DT 06-AUG-2003 (revised)  
 DT 22-FEB-2000 (first entry)  
 XX  
 DE D. pteronyssinus 98 kD mite allergen protein (map) Pderp98-509.  
 XX  
 XX Mite allergen protein; map; high molecular weight; HWM-map; allergy;  
 KM house dust mite; IGE; immunoglobulin E; allergen; mapB;  
 KM hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
 KM canine; veterinary; antibody; vaccine; immunisation.  
 XX  
 OS Dermatophagoides pteronyssinus.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..19  
 FT Protein /note= "Signal peptide"  
 FT /note= "Mature Pderp98-509"  
 XX  
 PN W09954349-A2.  
 PD  
 XX 28-OCT-1999.  
 PF  
 XX 16-APR-1999; 99WO-US008524.  
 PF  
 XX 17-APR-1998; 98US-00062013.



CC acid. The Der HMW-map protein is useful for eliciting an immune response  
 CC against Der HMW-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitizing a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig)E or Der HMW-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HMW-map polypeptides of the invention  
 XX  
 SQ Sequence 509 AA;

Query Match 86.5%; Score 83; DB 5; Length 509;  
 Best Local Similarity 88.9%; Pred. No. 1.2e-05;  
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 DAFEPHGILLTAIVSPGK 18  
 :|||||:|||||:|||||:  
 DB 185 EAFEPHGILLTAIVSPGK 202

RESULT 13  
 ABP72624  
 ID ABP72624 standard; protein; 527 AA.  
 XX  
 AC ABP72624;  
 XX  
 DT 23-OCT-2003 (revised)  
 DT 11-JUN-2003 (first entry)  
 XX  
 DE Prawn chitinase.  
 XX  
 KM Insecticide; pesticide; insect control; insect; toxin; chitinase; enzyme;  
 KM neuropeptide; transgenic plant; crop protection; prawn.  
 XX  
 OS Marsupinaeus japonicus.  
 OS  
 PN MO2003014150-A2.  
 XX  
 PD 20-FEB-2003.  
 XX  
 PF 06-AUG-2002; 2002WO-GB003598.  
 XX  
 PR 08-AUG-2001; 2001GB-00019274.  
 XX  
 XX (UYDU-) UNIT DURHAM.  
 PA (ENVJ-) DEPT ENVIRONMENT FOOD & RURAL AFFAIRS.  
 XX  
 PI Gatehouse JA, Fitches EC, Edwards JP;  
 XX  
 DR WPI; 2003-278469/27.  
 XX  
 PT Fusion protein useful for combating insect pests, comprises a  
 PT translocating moiety comprising a plant protein capable of acting as a  
 PT carrier to translocate toxic moiety inside plant pathogen, and a toxic  
 PT moiety.  
 XX  
 XX Claim 7; Fig 7; 51pp; English.  
 XX  
 CC The present sequence is that of prawn (Penaeus japonicus) chitinase. This  
 CC protein can be used in claimed fusion proteins of the invention  
 CC comprising a translocating moiety and a toxic moiety, where the  
 CC translocating moiety is a plant protein (e.g. a lectin) capable of acting  
 CC as a carrier to translocate the toxic moiety across the gut wall of a  
 CC plant pathogen, and the toxic moiety is an arthropod-derived peptide or  
 CC protein capable of causing deleterious effects on growth, development,  
 CC reproduction or mortality in pest insects. Suitable arthropod peptides  
 CC and proteins include allostastins, chitinase, diuretic hormone and their  
 CC metabolites and analogues. Polynucleotides encoding the fusion protein,

CC vectors, host cells and transgenic plants that are resistant to disease  
 CC are also provided. The fusion protein is target-specific, and resists  
 CC degradation in the insect gut. (Updated on 23-Oct-2003 to standardise OS  
 CC field)  
 XX  
 SQ Sequence 527 AA;

Query Match 65.6%; Score 63; DB 6; Length 527;  
 Best Local Similarity 80.0%; Pred. No. 0.04;  
 Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 AFEPHGILLTAIVSP 16  
 :|||||:|||||:|||||:  
 DB 301 AFEPHGILLTAIVSP 315

RESULT 14  
 ABB66399  
 ID ABB66399 standard; protein; 508 AA.  
 XX  
 AC ABB66399;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 25989.  
 DE  
 KM Drosophila; developmental biology; cell signalling; insecticide;  
 KM pharmaceutical.  
 KM  
 OS Drosophila melanogaster.  
 OS  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US009231.  
 XX  
 PR 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX  
 PA (PERE) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li FWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR N-PESDB; ABL10502.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX  
 PS Disclosure; SEQ ID NO 25989; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pcr\_sequences  
 XX

QY 2 AFEPHGILLTAIVSPGK 18  
 :|||||:|||||:|||||:  
 DB 294 AFEPHGILLTAIVSPGK 310

RESULT 15

ABP72634

ID ABP72634 standard; protein; 467 AA.

XX ABP72634;

AC 23-OCT-2003 (revised)

DT 11-JUN-2003 (first entry)

XX Prawn chitinase.

XX Insecticide; pesticide; insect control; insect; toxin; chitinase; enzyme;  
KM neurotoxic; transgenic plant; crop protection; prawn.

OS Marsupinaeus japonicus.

XX Key Location/Qualifiers

FT Active-site 118..125

PN W02003014150-A2.

PD 20-FEB-2003.

PF 06-AUG-2002; 2002WO-GB003598.

PR 08-AUG-2001; 2001GB-00019274.

PA (UYDU-) UNIV DURHAM.

PA (ENVI-) DEPT ENVIRONMENT FOOD & RURAL AFFAIRS.

PI Gatehouse JA, Fitches EC, Edwards JP;

DR WPI; 2003-278469/27.

XX Fusion protein useful for combating insect pests, comprises a  
PT translocating moiety comprising a plant protein capable of acting as a  
PT carrier to translocate toxic moiety inside plant pathogen, and a toxic  
PT moiety.

XX Claim 7; Fig 14; s1pp; English.

XX The present sequence is that of prawn (Penaeus japonicus) chitinase. This  
CC protein can be used in claimed fusion proteins of the invention  
CC comprising a translocating moiety and a toxic moiety, where the  
CC translocating moiety is a plant protein (e.g. a lectin) capable of acting  
CC as a carrier to translocate the toxic moiety across the gut wall of a  
CC plant pathogen, and the toxic moiety is an arthropod-derived peptide or  
CC protein capable of causing deleterious effects on growth, development,  
CC reproduction or mortality in pest insects. Suitable arthropod peptides  
CC and proteins include alaricins, chitinase, diuretic hormone and their  
CC metabolites and analogues. Polynucleotides encoding the fusion protein,  
CC vectors, host cells and transgenic plants that are resistant to disease  
CC are also provided. The fusion protein is target-specific, and resists  
CC degradation in the insect gut. (Updated on 23-OCT-2003 to standardise OS  
CC field)

XX Sequence 467 AA;

XX Query Match 59.4%; Score 57; DB 6; Length 467;

XX Best Local Similarity 70.6%; Pred. No. 0.39;

XX Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 AFEPRGYLTAAVSPK 18

DB 150 ALQPRGMLTAAVSNK 166

Search completed: March 22, 2004, 06:51:40  
Job time : 7.47304 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:59:34 ; Search time 3.68205 Seconds  
(without alignments)  
1265.926 Million cell updates/sec

Title: US-09-6662-293-6

Perfect score: 96  
Sequence: 1 DAFEPHGYLITAAVSPGK 18

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1049977 seqs, 258955339 residues

Total number of hits satisfying chosen parameters: 1049977

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCR\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCRUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	96	100.0	18	US-10-218-743-6	Sequence 6, Appl
2	96	100.0	535	US-10-218-743-21	Sequence 21, Appl
3	96	100.0	555	US-10-218-743-15	Sequence 15, Appl
4	96	100.0	555	US-10-218-743-18	Sequence 18, Appl
5	83	86.5	490	US-10-218-743-41	Sequence 41, Appl
6	83	86.5	509	US-10-218-743-35	Sequence 35, Appl
7	83	86.5	509	US-10-218-743-38	Sequence 38, Appl
8	47	49.0	375	US-10-369-493-17892	Sequence 17892, A
9	47	46.9	140	US-09-729-674-122	Sequence 122, App
10	44	45.8	502	US-09-844-948-5	Sequence 5, Appl
11	44	45.8	502	US-10-080-960-16	Sequence 16, Appl
12	43.5	45.3	391	US-10-422-536-17	Sequence 17, Appl
13	43	44.8	224	US-10-094-749-3125	Sequence 3125, Ap
14	43	44.8	305	US-10-424-599-237149	Sequence 237149,
15	43	44.8	318	US-10-231-253A-13	Sequence 13, Appl

16	43	44.8	655	US-10-369-493-23463	Sequence 23463, A
17	42.5	44.3	183	US-10-424-599-279804	Sequence 279804,
18	42.5	44.3	266	US-10-156-761-9331	Sequence 9331, Ap
19	42.5	44.3	687	US-10-198-070-111	Sequence 111, App
20	42	43.8	140	US-10-389-566-440	Sequence 440, App
21	42	43.8	302	US-10-092-900A-334	Sequence 334, App
22	42	43.8	318	US-10-369-493-9000	Sequence 9000, Ap
23	42	43.8	389	US-10-062-254-246	Sequence 246, App
24	42	43.8	434	US-10-424-599-159131	Sequence 159131,
25	42	43.8	481	US-10-369-493-20773	Sequence 20773, A
26	42	43.8	500	US-10-014-099F-57	Sequence 57, Appl
27	41	42.7	32	US-10-671-403-206	Sequence 206, App
28	41	42.7	32	US-10-670-844-206	Sequence 206, App
29	41	42.7	32	US-10-671-134-206	Sequence 206, App
30	41	42.7	32	US-10-670-844-206	Sequence 206, App
31	41	42.7	32	US-10-673-098-206	Sequence 206, App
32	41	42.7	167	US-10-104-047-3129	Sequence 3129, Ap
33	41	42.7	212	US-09-925-300-1577	Sequence 1577, Ap
34	41	42.7	372	US-10-282-122A-68332	Sequence 68332, A
35	41	42.7	377	US-10-425-114-64200	Sequence 64200, A
36	41	42.7	379	US-10-425-114-61241	Sequence 61241, A
37	41	42.7	379	US-10-425-114-61242	Sequence 61242, A
38	41	42.7	382	US-10-425-114-61243	Sequence 61243, A
39	41	42.7	382	US-10-425-114-62846	Sequence 62846, A
40	41	42.7	412	US-09-930-512-75	Sequence 75, Appl
41	41	42.7	479	US-10-282-122A-48638	Sequence 48638, A
42	41	42.7	3298	US-10-149-819-21	Sequence 21, Appl
43	41	42.7	3312	US-10-225-567A-656	Sequence 656, App
44	41	42.7	3312	US-10-038-854-67	Sequence 67, Appl
45	41	42.7	3313	US-09-737-149-29	Sequence 29, Appl

## ALIGNMENTS

RESULT 1  
US-10-218-743-6  
Sequence 6, Application US/10218743  
Publication No. US20030096779A1  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley Wu  
APPLICANT: Weber, Eric R.  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/10/218,743  
CURRENT FILING DATE: 2002-08-13  
PRIOR APPLICATION NUMBER: US/09/292,225  
PRIOR FILING DATE: 1999-04-15  
PRIOR APPLICATION NUMBER: 60/098,909  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/085,295  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/098,565  
PRIOR FILING DATE: 1998-04-17  
PRIOR APPLICATION NUMBER: 09/062,013  
PRIOR FILING DATE: 1998-04-17  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 6  
LENGTH: 18  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-10-218-743-6  
Query Match 100.0%; Score 96; DB 14; Length 18;  
Best Local Similarity 100.0%; Pred. No. 5.1e-09;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cy 1 DAFEPHGYLITAAVSPGK 18  
Db 1 DAFEPHGYLITAAVSPGK 18

RESULT 2  
US-10-218-743-21  
; Sequence 21, Application US/10218743  
; Publication No. US20030096779A1  
; GENERAL INFORMATION:  
; APPLICANT: McCall, Catherine A.  
; APPLICANT: Hunter, Shirley Wu  
; APPLICANT: Weber, Eric R.  
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: AL-2-C3  
; CURRENT APPLICATION NUMBER: US/10/218,743  
; CURRENT FILING DATE: 2002-08-13  
; PRIOR APPLICATION NUMBER: US/09/292,225  
; PRIOR FILING DATE: 1999-04-15  
; PRIOR APPLICATION NUMBER: 60/098,909  
; PRIOR FILING DATE: 1998-09-02  
; PRIOR APPLICATION NUMBER: 60/085,295  
; PRIOR FILING DATE: 1998-05-13  
; PRIOR APPLICATION NUMBER: 60/098,565  
; PRIOR FILING DATE: 1998-04-17  
; PRIOR APPLICATION NUMBER: 09/062,013  
; PRIOR FILING DATE: 1998-04-17  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 21  
; LENGTH: 536  
; TYPE: PRT  
; ORGANISM: Dermatophagoides farinae  
US-10-218-743-21

Query Match 100.0%; Score 96; DB 14; Length 536;  
Best Local Similarity 100.0%; Pred. No. 2.8e-07;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 DAFEPHGYLTTAAVSPGK 18  
|||  
Db 166 DAFEPHGYLTTAAVSPGK 183

RESULT 3  
US-10-218-743-15  
; Sequence 15, Application US/10218743  
; Publication No. US20030096779A1  
; GENERAL INFORMATION:  
; APPLICANT: McCall, Catherine A.  
; APPLICANT: Hunter, Shirley Wu  
; APPLICANT: Weber, Eric R.  
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: AL-2-C3  
; CURRENT APPLICATION NUMBER: US/10/218,743  
; CURRENT FILING DATE: 2002-08-13  
; PRIOR APPLICATION NUMBER: US/09/292,225  
; PRIOR FILING DATE: 1999-04-15  
; PRIOR APPLICATION NUMBER: 60/098,909  
; PRIOR FILING DATE: 1998-09-02  
; PRIOR APPLICATION NUMBER: 60/085,295  
; PRIOR FILING DATE: 1998-05-13  
; PRIOR APPLICATION NUMBER: 60/098,565  
; PRIOR FILING DATE: 1998-04-17  
; PRIOR APPLICATION NUMBER: 09/062,013  
; PRIOR FILING DATE: 1998-04-17  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 15  
; LENGTH: 555  
; TYPE: PRT  
; ORGANISM: Dermatophagoides farinae  
US-10-218-743-15

Query Match 100.0%; Score 96; DB 14; Length 555;  
Best Local Similarity 100.0%; Pred. No. 2.9e-07;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 DAFEPHGYLTTAAVSPGK 18  
|||  
Db 185 DAFEPHGYLTTAAVSPGK 202

RESULT 4  
US-10-218-743-18  
; Sequence 18, Application US/10218743  
; Publication No. US20030096779A1  
; GENERAL INFORMATION:  
; APPLICANT: McCall, Catherine A.  
; APPLICANT: Hunter, Shirley Wu  
; APPLICANT: Weber, Eric R.  
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: AL-2-C3  
; CURRENT APPLICATION NUMBER: US/10/218,743  
; CURRENT FILING DATE: 2002-08-13  
; PRIOR APPLICATION NUMBER: US/09/292,225  
; PRIOR FILING DATE: 1999-04-15  
; PRIOR APPLICATION NUMBER: 60/098,909  
; PRIOR FILING DATE: 1998-09-02  
; PRIOR APPLICATION NUMBER: 60/085,295  
; PRIOR FILING DATE: 1998-05-13  
; PRIOR APPLICATION NUMBER: 60/098,565  
; PRIOR FILING DATE: 1998-04-17  
; PRIOR APPLICATION NUMBER: 09/062,013  
; PRIOR FILING DATE: 1998-04-17  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 18  
; LENGTH: 555  
; TYPE: PRT  
; ORGANISM: Dermatophagoides farinae  
US-10-218-743-18

Query Match 100.0%; Score 96; DB 14; Length 555;  
Best Local Similarity 100.0%; Pred. No. 2.9e-07;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 DAFEPHGYLTTAAVSPGK 18  
|||  
Db 185 DAFEPHGYLTTAAVSPGK 202

RESULT 5  
US-10-218-743-41  
; Sequence 41, Application US/10218743  
; Publication No. US20030096779A1  
; GENERAL INFORMATION:  
; APPLICANT: McCall, Catherine A.  
; APPLICANT: Hunter, Shirley Wu  
; APPLICANT: Weber, Eric R.  
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: AL-2-C3  
; CURRENT APPLICATION NUMBER: US/10/218,743  
; CURRENT FILING DATE: 2002-08-13  
; PRIOR APPLICATION NUMBER: US/09/292,225  
; PRIOR FILING DATE: 1999-04-15  
; PRIOR APPLICATION NUMBER: 60/098,909  
; PRIOR FILING DATE: 1998-09-02  
; PRIOR APPLICATION NUMBER: 60/085,295  
; PRIOR FILING DATE: 1998-05-13  
; PRIOR APPLICATION NUMBER: 60/098,565  
; PRIOR FILING DATE: 1998-04-17  
; PRIOR APPLICATION NUMBER: 09/062,013  
; PRIOR FILING DATE: 1998-04-17  
; NUMBER OF SEQ ID NOS: 49

SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 41  
; LENGTH: 490  
; TYPE: PRT  
; ORGANISM: Dermatophagoides farinae  
US-10-218-743-41

Query Match 86.5%; Score 83; DB 14; Length 490;  
Best Local Similarity 88.9%; Pred. No. 3.7e-05;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 DAFEPHGYLITAAVSPGK 18  
:|||||  
Db 166 EAFEPFGYLLTAAVSPGK 183

RESULT 6  
US-10-218-743-35  
; Sequence 35, Application US/10218743  
; Publication No. US20030096779A1  
; GENERAL INFORMATION:  
; APPLICANT: McCall, Catherine A.  
; APPLICANT: Hunter, Shirley Wu  
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
; FILE REFERENCE: AL-2-C3  
; CURRENT APPLICATION NUMBER: US/10/218,743  
; CURRENT FILING DATE: 2002-08-13  
; PRIOR APPLICATION NUMBER: US/09/292,225  
; PRIOR FILING DATE: 1999-04-15  
; PRIOR APPLICATION NUMBER: 60/098,909  
; PRIOR FILING DATE: 1998-09-02  
; PRIOR APPLICATION NUMBER: 60/085,295  
; PRIOR FILING DATE: 1998-05-13  
; PRIOR APPLICATION NUMBER: 60/098,565  
; PRIOR FILING DATE: 1998-04-17  
; PRIOR APPLICATION NUMBER: 09/062,013  
; PRIOR FILING DATE: 1998-04-17  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 35  
; LENGTH: 509  
; TYPE: PRT  
; ORGANISM: Dermatophagoides farinae  
US-10-218-743-35

Query Match 86.5%; Score 83; DB 14; Length 509;  
Best Local Similarity 88.9%; Pred. No. 3.7e-05;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 DAFEPHGYLITAAVSPGK 18  
:|||||  
Db 185 EAFEPFGYLLTAAVSPGK 202

RESULT 7  
US-10-218-743-38  
; Sequence 38, Application US/10218743  
; Publication No. US20030096779A1  
; GENERAL INFORMATION:  
; APPLICANT: McCall, Catherine A.  
; APPLICANT: Hunter, Shirley Wu  
; APPLICANT: Weber, Eric R.  
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
; FILE REFERENCE: AL-2-C3  
; CURRENT APPLICATION NUMBER: US/10/218,743  
; CURRENT FILING DATE: 2002-08-13  
; PRIOR APPLICATION NUMBER: US/09/292,225  
; PRIOR FILING DATE: 1999-04-15  
; PRIOR APPLICATION NUMBER: 60/098,909  
; PRIOR FILING DATE: 1998-09-02

; PRIOR APPLICATION NUMBER: 60/085,295  
; PRIOR FILING DATE: 1998-05-13  
; PRIOR APPLICATION NUMBER: 60/098,565  
; PRIOR FILING DATE: 1998-04-17  
; PRIOR APPLICATION NUMBER: 09/062,013  
; PRIOR FILING DATE: 1998-04-17  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 38  
; LENGTH: 509  
; TYPE: PRT  
; ORGANISM: Dermatophagoides farinae  
US-10-218-743-38

Query Match 86.5%; Score 83; DB 14; Length 509;  
Best Local Similarity 88.9%; Pred. No. 3.7e-05;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 DAFEPHGYLITAAVSPGK 18  
:|||||  
Db 185 EAFEPFGYLLTAAVSPGK 202

RESULT 8  
US-10-369-493-17892  
; Sequence 17892, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 17892  
; LENGTH: 375  
; TYPE: PRT  
; ORGANISM: SPRINGCOMONS  
US-10-369-493-17892

Query Match 49.0%; Score 47; DB 15; Length 375;  
Best Local Similarity 58.8%; Pred. No. 24;  
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 DAFEPHGYLITAAVSPG 17  
:|||||  
Db 215 NAFHPYGVLLRAADVGV 231

RESULT 9  
US-09-729-674-122  
; Sequence 122, Application US/09729674  
; Patent No. US20010039335A1  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John M.  
; APPLICANT: Lavallee, Edward R.  
; APPLICANT: Collins-Racie, Lisa A.  
; APPLICANT: Evans, Cheryl  
; APPLICANT: Merberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Steindinger II, Robert J.  
; APPLICANT: Spaulding, Vikki  
; APPLICANT: Wong, Gordon G.  
; APPLICANT: Clark, Hilary

APPLICANT: Pechtel, Kim  
APPLICANT: Genetics Institute, Inc.  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
FILE REFERENCE: 6055-64X  
CURRENT APPLICATION NUMBER: US/09/729,674  
CURRENT FILING DATE: 2000-12-04  
PRIOR APPLICATION NUMBER: 09/539,330  
PRIOR FILING DATE: 2000-03-30  
NUMBER OF SEQ ID NOS: 283  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 122  
LENGTH: 140  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-729-674-122

Query Match 46.9%; Score 45; DB 9; Length 140;  
Best Local Similarity 50.0%; Pred. No. 17;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 EPHGYLTTAAVSPG 17  
DB 51 QPHAFLLSGPRSPG 64

RESULT 10  
US-09-844-948-5  
Sequence 5, Application US/09844948  
Publication No. US20030119161A1  
GENERAL INFORMATION:  
APPLICANT: Meyers, Rachel A.  
APPLICANT: Williamson, Mark  
TITLE OF INVENTION: MEMBER AND USES THEREFOR  
FILE REFERENCE: 10448-048001  
CURRENT APPLICATION NUMBER: US/09/844,948  
CURRENT FILING DATE: 2001-04-24  
PRIOR APPLICATION NUMBER: US 60/200,604  
PRIOR FILING DATE: 2000-04-28  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5  
LENGTH: 502  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: consensus sequence  
US-09-844-948-5

Query Match 45.8%; Score 44; DB 10; Length 502;  
Best Local Similarity 71.4%; Pred. No. 1.1e+02;  
Matches 10; Conservative 1; Mismatches 1; Indels 2; Gaps 1;

QY 1 DAFEPHGYL--LTA 12  
DB 229 DAFHPGYLLEPLTA 242

RESULT 11  
US-10-080-960-16  
Sequence 16, Application US/10080960  
Publication No. US20020197695A1  
GENERAL INFORMATION:  
APPLICANT: Millennium Pharmaceuticals, Inc.  
APPLICANT: Glucksmann, Maria  
APPLICANT: Meyers, Rachel  
TITLE OF INVENTION: 80090, 52874, 52880, 63497, AND 33425  
TITLE OF INVENTION: METHODS AND COMPOSITIONS OF HUMAN PROTEINS AND USES THEREOF  
FILE REFERENCE: 38155-2004.00  
CURRENT APPLICATION NUMBER: US/10/080,960  
CURRENT FILING DATE: 2001-10-19  
PRIOR APPLICATION NUMBER: US 60/242,040  
PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: US 60/242,038  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: US 60/241,992  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: US 60/242,637  
PRIOR FILING DATE: 2000-10-23  
NUMBER OF SEQ ID NOS: 37  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 16  
LENGTH: 502  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Consensus amino acid sequence  
US-10-080-960-16

Query Match 45.8%; Score 44; DB 13; Length 502;  
Best Local Similarity 71.4%; Pred. No. 1.1e+02;  
Matches 10; Conservative 1; Mismatches 1; Indels 2; Gaps 1;

QY 1 DAFEPHGYL--LTA 12  
DB 229 DAFHPGYLLEPLTA 242

RESULT 12  
US-10-422-536-17  
Sequence 17, Application US/10422536  
Publication No. US20040014100A1  
GENERAL INFORMATION:  
APPLICANT: Kinseila, Todd  
APPLICANT: Lorens, James  
APPLICANT: Pray, Todd  
TITLE OF INVENTION: IN VITRO PRODUCTION OF CYCLIC PEPTIDES FOR INHIBITING  
TITLE OF INVENTION: PROTEIN-PROTEIN INTERACTION  
FILE REFERENCE: A-71433-1/AMP/CYO  
CURRENT APPLICATION NUMBER: US/10/422,536  
CURRENT FILING DATE: 2003-04-23  
PRIOR APPLICATION NUMBER: US 60/187,130  
PRIOR FILING DATE: 2000-03-06  
PRIOR APPLICATION NUMBER: US 09/800,770  
PRIOR FILING DATE: 2001-03-06  
PRIOR APPLICATION NUMBER: US 10/232,758  
PRIOR FILING DATE: 2002-08-30  
NUMBER OF SEQ ID NOS: 168  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 17  
LENGTH: 391  
TYPE: PRT  
ORGANISM: Thermococcus hydrothermalis  
US-10-422-536-17

Query Match 45.3%; Score 43.5; DB 15; Length 391;  
Best Local Similarity 47.6%; Pred. No. .96;  
Matches 10; Conservative 2; Mismatches 6; Indels 3; Gaps 1;

QY 1 DAFEPH--GYLTTAAVSPG 18  
DB 84 DVTEDHSLIGYMWTSVKPKG 104

RESULT 13  
US-10-094-749-3125  
Sequence 3125, Application US/10094749  
Publication No. US20030219741A1  
GENERAL INFORMATION:  
APPLICANT: ISOGAI, TAKAO  
APPLICANT: SUGIYAMA, TOMOYASU  
APPLICANT: OTSUKI, TETSUJI  
APPLICANT: WAKAMATSU, AI  
APPLICANT: SATO, HIROYUKI  
APPLICANT: ISHII, SHIZUKO



APPLICANT: YAMAMOTO, JUN-ICHI  
APPLICANT: ISONO, YUKO  
APPLICANT: HIO, YURI  
APPLICANT: OTSUKA, KAORU  
APPLICANT: NAGAI, KEIICHI  
APPLICANT: IRIE, RYOTARO  
APPLICANT: TAMECHIKA, ICHIRO  
APPLICANT: SEKI, NAOHITO  
APPLICANT: YOSHIKAWA, TSUTOMU  
APPLICANT: OTSUKA, MOTYUKI  
APPLICANT: NAGAHARI, KENJI  
APPLICANT: MASUHO, YASUHIRO  
TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA  
FILE REFERENCE: 094335/0160  
CURRENT APPLICATION NUMBER: US/10/094,749  
CURRENT FILING DATE: 2002-03-12  
PRIOR APPLICATION NUMBER: 60/350,435  
PRIOR FILING DATE: 2002-01-24  
PRIOR APPLICATION NUMBER: JP 2001-328381  
PRIOR FILING DATE: 2001-09-14  
NUMBER OF SEQ ID NOS: 3381  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3125  
LENGTH: 224  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-094-749-3125

Query Match 44.8%; Score 43; DB 15; Length 224;  
Best Local Similarity 57.1%; Pred. No. 62;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 5 PHGVLTAAVSPK 18  
DB 136 PHGAVLAPTPPGK 149

RESULT 14  
US-10-424-599-237149  
Sequence 237149, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 237149  
LENGTH: 305  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(305)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_56170C.1.pep  
US-10-424-599-237149

Query Match 44.8%; Score 43; DB 12; Length 305;  
Best Local Similarity 56.2%; Pred. No. 88;  
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 AFEPHGYLTAAVSPG 17  
DB 32 SFLTSGYVLTATGSPG 47

RESULT 15  
US-10-291-253A-13  
Sequence 13, Application US/10291253A  
Publication No. US20030150017A1  
GENERAL INFORMATION:  
APPLICANT: Botella, Jose  
APPLICANT: Graham, Michael  
APPLICANT: Fairbairn, David  
TITLE OF INVENTION: A Method for Facilitating Pathogen Resistance  
FILE REFERENCE: nematode  
CURRENT APPLICATION NUMBER: US/10/291,253A  
CURRENT FILING DATE: 2003-03-31  
PRIOR APPLICATION NUMBER: PR8706  
PRIOR FILING DATE: 2001-11-07  
PRIOR APPLICATION NUMBER: PR8802  
PRIOR FILING DATE: 2001-11-12  
PRIOR APPLICATION NUMBER: US60/341404  
PRIOR FILING DATE: 2001-12-14  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 13  
LENGTH: 318  
TYPE: PRT  
ORGANISM: Nematode  
US-10-291-253A-13

Query Match 44.8%; Score 43; DB 14; Length 318;  
Best Local Similarity 60.0%; Pred. No. 92;  
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 AFEPHGYLTAAVSP 16  
DB 241 AFSFNRVWLCAVGP 255

Search completed: March 22, 2004, 07:45:48  
Job time : 3.68205 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: March 22, 2004, 06:42:54 ; Search time 1.45081 Seconds

(without alignments)  
640.518 Million cell updates/sec

Title: US-09-662-293-6

Perfect score: 96

Sequence: 1 DAFPHGYLITAAVSPGK 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Issued Patents AA: \*  
1: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/prodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/prodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/prodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/prodata/2/iaa/backfill1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	96	100.0	18	US-09-292-225-6	Sequence 6, Appl1
2	96	100.0	536	US-09-292-225-21	Sequence 21, Appl1
3	96	100.0	555	US-09-292-225-15	Sequence 15, Appl1
4	96	100.0	555	US-09-292-225-18	Sequence 18, Appl1
5	83	86.5	490	US-09-292-225-41	Sequence 41, Appl1
6	83	86.5	509	US-09-292-225-35	Sequence 35, Appl1
7	83	86.5	509	US-09-292-225-38	Sequence 38, Appl1
8	45	46.9	309	US-09-489-039A-13534	Sequence 13534, A
9	43.5	45.3	391	US-09-800-170-17	Sequence 17, Appl1
10	43	44.8	442	US-09-052-778-2	Sequence 2, Appl1
11	42	43.8	207	US-09-489-039A-9124	Sequence 9124, Ap
12	41.5	43.2	459	US-09-489-039A-12641	Sequence 12641, A
13	40	41.7	292	US-09-252-991A-19532	Sequence 19532, A
14	40	41.7	354	US-09-489-039A-8658	Sequence 8658, Ap
15	40	41.7	583	US-09-489-039A-113918	Sequence 113918, A
16	39	40.6	31	US-08-190-802A-104	Sequence 104, App
17	39	40.6	31	US-08-477-346-104	Sequence 104, App
18	39	40.6	31	US-08-473-089-104	Sequence 104, App
19	39	40.6	31	US-08-487-072A-104	Sequence 104, App
20	39	40.6	105	US-09-134-000C-3563	Sequence 3563, Ap
21	39	40.6	280	US-09-323-998E-37	Sequence 37, Appl
22	39	40.6	283	US-09-266-965-109	Sequence 109, Appl
23	39	40.6	318	US-08-150-802A-33	Sequence 33, Appl
24	39	40.6	318	US-08-477-346-33	Sequence 33, Appl
25	39	40.6	318	US-08-473-089-33	Sequence 33, Appl
26	39	40.6	318	US-08-487-072A-33	Sequence 33, Appl
27	39	40.6	350	US-08-828-922-1	Sequence 1, Appl1

28	39	40.6	410	US-09-540-236-3584	Sequence 3584, Ap
29	39	40.6	414	US-09-252-991A-31156	Sequence 31156, A
30	39	40.6	422	US-09-328-352-1491	Sequence 7491, Ap
31	39	40.6	513	US-09-252-991A-19670	Sequence 19670, A
32	39	40.6	713	US-09-198-452A-640	Sequence 640, App
33	39	40.6	938	US-09-637-145-2	Sequence 2, Appl1
34	39	40.6	981	US-09-252-991A-16920	Sequence 16920, A
35	39	40.6	1179	US-09-252-991A-17895	Sequence 17895, A
36	38.5	40.1	127	US-09-673-395A-595	Sequence 595, App
37	38.5	40.1	526	US-09-252-991A-20120	Sequence 20120, A
38	38.5	40.1	666	US-08-083-590A-17	Sequence 17, Appl
39	38.5	40.1	666	US-08-346-128-16	Sequence 36, Appl
40	38.5	40.1	666	US-08-532-384-17	Sequence 17, Appl
41	38.5	40.1	681	US-08-083-590A-18	Sequence 18, Appl
42	38.5	40.1	681	US-08-346-128-17	Sequence 37, Appl
43	38.5	40.1	681	US-08-532-384-18	Sequence 18, Appl
44	38.5	40.1	1068	US-08-537-210A-2	Sequence 2, Appl1
45	38.5	40.1	1068	US-09-113-825-2	Sequence 2, Appl1

## ALIGNMENTS

```
RESULT 1
US-09-292-225-6
; Sequence 6, Application US/09292225
; Patent No. 6455686
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; EARLIER FILING DATE: 1998-04-15
; EARLIER APPLICATION NUMBER: 60/098,909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085,295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098,565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062,013
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; TYPE: PRT
; LENGTH: 18
; ORGANISM: Dermatophagoides farinae
US-09-292-225-6
Query Match 100.0%; Score 96; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 3.3e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DAFPHGYLITAAVSPGK 18
DB 1 DAFPHGYLITAAVSPGK 18
RESULT 2
US-09-292-225-21
; Sequence 21, Application US/09292225
; Patent No. 6455686
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
```

```

; CURRENT FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: 60/098, 909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085, 295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098, 565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062, 013
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 21
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
; US-09-292-225-21

Query Match      100.0%; Score 96; DB 4; Length 536;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DAFEPHGILTAAVSPGK 18
      |||||
Db      166 DAFEPHGILTAAVSPGK 183

RESULT 3
US-09-292-225-15
; Sequence 15, Application US/09292225
; Patent No. 6455686
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; CURRENT FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: 60/098, 909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085, 295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098, 565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062, 013
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 15
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
; US-09-292-225-15

Query Match      100.0%; Score 96; DB 4; Length 555;
Best Local Similarity 100.0%; Pred. No. 2.1e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DAFEPHGILTAAVSPGK 18
      |||||
Db      185 DAFEPHGILTAAVSPGK 202

RESULT 4
US-09-292-225-18
; Sequence 18, Application US/09292225
; Patent No. 6455686
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; US-09-292-225-18
```

```

; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; CURRENT FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: 60/098, 909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085, 295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098, 565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062, 013
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 18
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
; US-09-292-225-18

Query Match      100.0%; Score 96; DB 4; Length 555;
Best Local Similarity 100.0%; Pred. No. 2.1e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DAFEPHGILTAAVSPGK 18
      |||||
Db      185 DAFEPHGILTAAVSPGK 202

RESULT 5
US-09-292-225-41
; Sequence 41, Application US/09292225
; Patent No. 6455686
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; CURRENT FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: 60/098, 909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085, 295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098, 565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062, 013
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 41
; LENGTH: 490
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
; US-09-292-225-41

Query Match      86.5%; Score 83; DB 4; Length 490;
Best Local Similarity 88.9%; Pred. No. 3.2e-06;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 DAFEPHGILTAAVSPGK 18
      :|||
Db      166 EAFEPHGILTAAVSPGK 183

RESULT 6
US-09-292-225-35
; Sequence 35, Application US/09292225
; Patent No. 6455686
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; US-09-292-225-35
```

```

; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AI-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; EARLIER FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: 60/098,909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085,295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098,565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062,013
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Dermacophagoides farinae
US-09-292-225-35

Query Match      86.5%; Score 83; DB 4; Length 509;
Best Local Similarity 88.9%; Pred. No. 3,4e-06;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 DAFPHGYLTLTAAVSPGK 18
;|||||
Db      185 EAFEPFGYLTAAVSPGK 202

RESULT 7
US-09-292-225-38
; Sequence 38, Application US/09292225
; Patent No. 6455686
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AI-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; EARLIER FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: 60/098,909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085,295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098,565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062,013
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Dermacophagoides farinae
US-09-292-225-38

Query Match      86.5%; Score 83; DB 4; Length 509;
Best Local Similarity 88.9%; Pred. No. 3,4e-06;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 DAFPHGYLTLTAAVSPGK 18
;|||||
Db      185 EAFEPFGYLTAAVSPGK 202

RESULT 8
US-09-489-039A-13534
; Sequence 13534, Application US/09489039A
```

```

; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Bretton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; EARLIER FILING DATE: 2000-01-27
; EARLIER APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13534
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13534

Query Match      46.9%; Score 45; DB 4; Length 309;
Best Local Similarity 56.2%; Pred. No. 7.5;
Matches 9; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY      1 DAFPHGYLTLTAAVSP 16
;|||||
Db      61 DGFAPHGNLTAAVSLP 76

RESULT 9
US-09-800-170-17
; Sequence 17, Application US/09800170
; Patent No. 6481667
; GENERAL INFORMATION:
; APPLICANT: Kinsella, Todd
; TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES
; FILE REFERENCE: A-68614-1/DJB/RMS/RMK
; CURRENT APPLICATION NUMBER: US/09/800,170
; EARLIER FILING DATE: 2001-08-28
; EARLIER APPLICATION NUMBER: US 60/187,130
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Thermococcus hydrothermalis
US-09-800-170-17

Query Match      45.3%; Score 43.5; DB 4; Length 391;
Best Local Similarity 47.6%; Pred. No. 18;
Matches 10; Conservative 2; Mismatches 6; Indels 3; Gaps 1;

QY      1 DAFEPH---GYLTLTAAVSPGK 18
;|||||
Db      84 DVTEDHSLIGYNTSKVKGK 104

RESULT 10
US-09-052-778-2
; Sequence 2, Application US/09052778A
; Patent No. 6060590
; GENERAL INFORMATION:
; APPLICANT: Bryant, Peter J.
; APPLICANT: Kawamura, Kazuo
; TITLE OF INVENTION: CHITINASE RELATED PROTEINS AND METHODS
; FILE REFERENCE: 07306/015001
; CURRENT APPLICATION NUMBER: US/09/052,778A
; EARLIER FILING DATE: 1998-03-31
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
```

US-09-052-778-2

Query Match 44.8%; Score 43; DB 3; Length 442;  
Best Local Similarity 50.0%; Pred. No. 26;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 DAFEPHYLITAAVSP 16

Db 206 NAFRPGYILGLSVLP 221

RESULT 11

US-09-489-039A-9124  
Sequence 9124, Application US/09489039A  
Patent No. 6610836  
GENERAL INFORMATION:  
APPLICANT: Gary Breton et. al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
FILE REFERENCE: 2709.2004001  
CURRENT APPLICATION NUMBER: US/09/489,039A  
CURRENT FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: US 60/117,747  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 14342  
SEQ ID NO 9124  
LENGTH: 207  
TYPE: PRT  
ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-9124

Query Match 43.8%; Score 42; DB 4; Length 207;  
Best Local Similarity 56.2%; Pred. No. 15;  
Matches 9; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 AFEPHYLITAAVSP 17

Db 82 AAEPPHAAQAAVAAPFG 97

RESULT 12  
US-09-489-039A-12641  
Sequence 12641, Application US/09489039A  
Patent No. 6610836  
GENERAL INFORMATION:  
APPLICANT: Gary Breton et. al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
FILE REFERENCE: 2709.2004001  
CURRENT APPLICATION NUMBER: US/09/489,039A  
CURRENT FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: US 60/117,747  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 14342  
SEQ ID NO 12641  
LENGTH: 459  
TYPE: PRT  
ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-12641

Query Match 43.2%; Score 41.5; DB 4; Length 459;  
Best Local Similarity 42.9%; Pred. No. 49;  
Matches 9; Conservative 3; Mismatches 2; Indels 7; Gaps 1;

QY 4 EPHGY-----LITAAVSP 17

Db 229 QPRGYIAFKPGILTSVQPG 249

RESULT 13  
US-09-252-991A-19532  
Sequence 19532, Application US/09252991A  
Patent No. 6551795

GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 19532  
LENGTH: 292  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-19532

Query Match 41.7%; Score 40; DB 4; Length 292;  
Best Local Similarity 57.1%; Pred. No. 52;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 5 PEGYLLTAAVSPGK 18

Db 245 PPGYPTGLVSPGR 258

RESULT 14  
US-09-489-039A-8658  
Sequence 8658, Application US/09489039A  
Patent No. 6610836  
GENERAL INFORMATION:  
APPLICANT: Gary Breton et. al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
FILE REFERENCE: 2709.2004001  
CURRENT APPLICATION NUMBER: US/09/489,039A  
CURRENT FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: US 60/117,747  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 14342  
SEQ ID NO 8658  
LENGTH: 354  
TYPE: PRT  
ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-8658

Query Match 41.7%; Score 40; DB 4; Length 354;  
Best Local Similarity 44.4%; Pred. No. 65;  
Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 DAFEPHYLITAAVSPGK 18

Db 82 DAIDPHADLVVAISQSGK 99

RESULT 15  
US-09-489-039A-13918  
Sequence 13918, Application US/09489039A  
Patent No. 6610836  
GENERAL INFORMATION:  
APPLICANT: Gary Breton et. al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
FILE REFERENCE: 2709.2004001  
CURRENT APPLICATION NUMBER: US/09/489,039A  
CURRENT FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: US 60/117,747  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 14342  
SEQ ID NO 13918  
LENGTH: 583  
TYPE: PRT

ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-13918

Query Match 41.7%; Score 40; DB 4; Length 583;  
Best Local Similarity: 80.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 8 YLTPAAVSPG 17  
Db 547 YLTPAAISAG 556

Search completed: March 22, 2004, 07:03:55  
Job time : 1.45081 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:40:28 ; Search time 0.84547 Seconds  
(without alignments)  
1479.047 Million cell updates/sec

Title: US-09-662-293-7  
Perfect score: 64  
Sequence: 1 DKONYALVRELK 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: PIR78:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR4:\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	48	75.0	1484 2 T29275	hypothetical prote
2	46	71.9	617 2 T15408	hypothetical prote
3	44	68.8	435 2 T24074	hypothetical prote
4	42	65.6	399 2 T19115	hypothetical prote
5	42	65.6	460 2 T19111	hypothetical prote
6	42	65.6	465 2 T19113	hypothetical prote
7	42	65.6	522 2 T28113	hypothetical prote
8	42	65.6	633 2 T24898	hypothetical prote
9	42	65.6	831 2 T00323	chitinase (EC 3.2.
10	41	64.1	462 2 S76693	hypothetical prote
11	41	64.1	756 2 AB1088	chitinase B homolo
12	41	64.1	756 2 AB1452	chitinase B homolo
13	41	64.1	2708 2 T09079	probable chloroqui
14	41	64.1	2819 2 T09080	probable chloroqui
15	40	62.5	84 2 F75551	acyl-CoA-binding p
16	40	62.5	379 2 S73517	probable mannitol-
17	40	62.5	383 2 A49562	carililage glycopro
18	40	62.5	403 2 H70324	hypothetical prote
19	39	60.9	248 2 E84089	hypothetical prote
20	39	60.9	509 1 D69884	conserved hypothet
21	38	59.4	129 2 PQ0124	major merizolite su
22	38	59.4	129 2 PQ0123	major merizolite su
23	38	59.4	129 2 PQ0121	major merizolite su
24	38	59.4	138 2 PQ0125	major merizolite su
25	38	59.4	168 2 A25522	major merizolite su
26	38	59.4	594 2 E96667	unknown protein, 6
27	38	59.4	1060 2 S06286	major merizolite su
28	38	59.4	1701 2 A54498	major merizolite su
29	38	59.4	1701 2 A26868	major merizolite su

30	37	57.8	407 2 C64250	tyrosine-tRNA liga
31	37	57.8	731 2 D36808	hypothetical prote
32	37	57.8	764 2 T25012	hypothetical prote
33	37	57.8	809 2 B70692	hypothetical prote
34	37	57.8	944 2 S66870	DNA protein - yea
35	37	57.8	964 2 D59404	plectin isoform pl
36	37	57.8	4574 2 G02520	plectin - human
37	37	57.8	4684 2 A59404	plectin (imported)
38	37	57.8	4687 1 A39638	plectin - rat
39	36	56.2	165 2 C83415	hypothetical prote
40	36	56.2	177 2 B70190	conserved hypothet
41	36	56.2	183 2 F31090	H+-transporting tw
42	36	56.2	183 2 AF1807	ATP synthase chain
43	36	56.2	217 2 T26669	hypothetical prote
44	36	56.2	262 2 F70375	cell division inh
45	36	56.2	371 2 D70357	conserved hypothet

ALIGNMENTS

RESULT 1  
T29275  
hypothetical protein T01C4.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T29275  
R:Graves, T.; Wohldmann, P.  
Submitted to the EMBL Data Library, September 1996  
A:Description: The sequence of C. elegans cosmid T01C4.  
A:Reference number: Z20599  
A:Accession: T29275  
A:Status: Preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-1484 <GRA>  
A:Cross-references: EMBL:U70858; PIDN:AA09177.1; GSPDB:GN00023; CESP:T01C4.1  
A:Experimental source: strain Bristol N2; clone T01C4  
C:Genetics:  
A:Gene: CESP:T01C4.1  
A:Map position: 5  
A:Introns: 60/1; 179/1; 219/3; 392/3; 481/3; 549/2; 594/2; 736/3; 843/3; 1056/2; 1444/3

Query Match 75.0%; Score 48; DB 2; Length 1484;  
Best Local Similarity 61.5%; Pred. No. 2.3;  
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKONYALVRELK 13  
DB 1087 DKONYALVRELK 1099

RESULT 2  
T15408  
hypothetical protein C04F6.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T15408  
R:Nhan, M.  
Submitted to the EMBL Data Library, December 1995  
A:Description: The sequence of C. elegans cosmid C04F6.  
A:Reference number: Z18346  
A:Accession: T15408  
A:Status: Preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-617 <NHA>  
A:Cross-references: EMBL:U42835; NID:G1125760; PID:G1125762; PIDN:AAA83586.1; CESP:C04F6.  
A:Gene: CESP:C04F6.3  
A:Introns: 28/1; 66/2; 504/1

Query Match 71.9%; Score 46; DB 2; Length 617;  
Best Local Similarity 69.2%; Pred. No. 2.2;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DKONYTALVRELK 13  
|:|:|:|:|:  
Db 186 DMANTVALVRELK 198

## RESULT 3

T24074

hypothetical protein R09D1.3 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C/Accession: T24074

R/Matthews, P.

submitted to the EMBL Data Library, March 1996

A/Reference number: Z19838

A/Accession: T24074

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-435 &lt;MIL&gt;

A/Cross-references: EMBL:Z70035; PIDN:CAA93863.1; GSPDB:GN00020; CESP:R09D1.3

A/Experimental source: clone R09D1

C/Genetics:

A/Gene: CESP:R09D1.3

A/Map position: 2

A/Introns: 12/2; 71/1; 165/2

Query Match 68.8%; Score 44; DB 2; Length 435;

Best Local Similarity 53.8%; Pred. No. 3.6;

Matches 7; Conservative 3; Mismatches 0; Gaps 0;

Qy 1 DKONYTALVRELK 13  
|:|:|:|:|:  
Db 194 DKSNYATLIREIR 206

## RESULT 4

T19115

hypothetical protein C08H9.14 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C/Accession: T19115

R/Lloyd, C.

submitted to the EMBL Data Library, October 1995

A/Reference number: Z19074

A/Accession: T19115

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-359 &lt;MIL&gt;

A/Cross-references: EMBL:Z54342; PIDN:CAA91153.1; GSPDB:GN00020; CESP:C08H9.14

A/Experimental source: clone C08H9

C/Genetics:

A/Gene: CESP:C08H9.14

A/Map position: 2

A/Introns: 39/1; 134/2

Query Match 65.6%; Score 42; DB 2; Length 399;

Best Local Similarity 53.8%; Pred. No. 7.7;

Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 DKONYTALVRELK 13  
|:|:|:|:|:  
Db 163 DQSNYLIIFIREIR 175

## RESULT 5

T19111

hypothetical protein C08H9.12 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C/Accession: T19111

R/Lloyd, C.

submitted to the EMBL Data Library, October 1995

A/Reference number: Z19074

A/Accession: T19111  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-460 <MIL>  
A/Cross-references: EMBL:Z54342; PIDN:CAA91149.1; GSPDB:GN00020; CESP:C08H9.12  
A/Experimental source: clone C08H9  
C/Genetics:  
A/Gene: CESP:C08H9.12  
A/Map position: 2  
A/Introns: 19/2; 69/1; 99/1

Query Match 65.6%; Score 42; DB 2; Length 460;

Best Local Similarity 53.8%; Pred. No. 8.9;

Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 DKONYTALVRELK 13  
|:|:|:|:|:  
Db 229 DENNYLMFIREIR 241

## RESULT 6

T19113

hypothetical protein C08H9.4 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C/Accession: T19113

R/Lloyd, C.

submitted to the EMBL Data Library, October 1995

A/Reference number: Z19074

A/Accession: T19113

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-465 &lt;MIL&gt;

A/Cross-references: EMBL:Z54342; PIDN:CAA91151.1; GSPDB:GN00020; CESP:C08H9.4

A/Experimental source: clone C08H9

C/Genetics:

A/Gene: CESP:C08H9.4

A/Map position: 2

A/Introns: 16/2; 66/1; 104/1

Query Match 65.6%; Score 42; DB 2; Length 465;  
Best Local Similarity 53.8%; Pred. No. 9;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 DKONYTALVRELK 13  
|:|:|:|:|:  
Db 234 DKNNYLMFIREIR 246

## RESULT 7

T28113

hypothetical protein ZK938.6 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999

C/Accession: T28113

R/Lloyd, C.

submitted to the EMBL Data Library, June 1995

A/Reference number: Z20471

A/Accession: T28113

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-522 &lt;MIL&gt;

A/Cross-references: EMBL:Z49913; PIDN:CAA90145.1; GSPDB:GN00020; CESP:ZK938.6

A/Experimental source: clone ZK938

C/Genetics:

A/Gene: CESP:ZK938.6

A/Map position: 2

A/Introns: 161/1

Query Match 65.6%; Score 42; DB 2; Length 522;  
Best Local Similarity 53.8%; Pred. No. 10;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;



Oy 1 DKONYALVRELK 13  
 Db 291 DENNYLMFIRELR 303

## RESULT 8

T24898  
 hypothetical protein T13H5.3 - *Caenorhabditis elegans*

C/Species: *Caenorhabditis elegans*  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C/Accession: T24898  
 R/ighting, J.  
 submitted to the EMBL Data Library, October 1995

A/Reference number: Z19950  
 A/Accession: T24898  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-633 <WIL>  
 A/Cross-references: EMBL:Z66524; PIDD:CA91419.1; GSPDB:GN00020; CESP:T13H5.3  
 A/Experimental source: clone T13H5  
 C/Genetics:  
 A/Gene: CESP:T13H5.3  
 A/Map position: 2  
 A/Intons: 27/1; 57/3; 93/3; 126/2; 251/2; 274/3; 457/3; 578/1

Query Match 65.6%; Score 42; DB 2; Length 633;  
 Best Local Similarity 53.8%; Pred. No. 12;  
 Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Oy 1 DKONYALVRELK 13  
 Db 164 DRRNYVIMRELR 176

## RESULT 9

T00323  
 chitinase (EC 3.2.1.14) B - *Clostridium paraputrificum*

N/Alternate names: 1,4-beta-poly-N-acetylglucosaminidase; chitodextrinase  
 C/Species: *Clostridium paraputrificum*  
 C/Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 25-Aug-2003  
 C/Accession: T00323  
 R/Omiya, K.; Sakka, K.; Kimura, T.; Karita, S.; Morimoto, K.  
 J. Bacteriol. 179, 7306-7314, 1997  
 A/Title: Cloning, sequencing, and expression of the gene encoding *Clostridium paraputrificum*  
 A/Reference number: Z14138; MUID:98055843; PMID:9393694  
 A/Accession: T00323  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-831 <OHM>  
 A/Cross-references: EMBL:AB001874; NID:g2696016; PIDD:BAA23796.1; PID:g2696017  
 A/Experimental source: strain M21  
 C/Genetics:  
 A/Gene: ch1B  
 C/Superfamily: chitinase with C-terminal chitin-binding domain  
 C/Keywords: glycosidase; hydrolase

Query Match 65.6%; Score 42; DB 2; Length 831;  
 Best Local Similarity 53.8%; Pred. No. 16;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Oy 1 DKONYALVRELK 13  
 Db 210 DKONYITLQDK 222

## RESULT 10

S76693  
 hypothetical protein - *Synechocystis* sp. (strain PCC 6803)

C/Species: *Synechocystis* sp.  
 A/Variety: PCC 6803  
 C/Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
 C/Accession: S76693  
 R/Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.

O, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Matanabe, A.; Yamada, M.; Yasuda  
 DNA Res. 3, 109-136, 1996  
 A/Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*  
 S.

A/Reference number: S74322; MUID:97061201; PMID:8905231  
 A/Accession: S76693  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-462 <KAN>  
 A/Cross-references: EMBL:D64004; GB:AB001339; NID:g1001701; PIDD:BAA10637.1; PID:d101128  
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 64.1%; Score 41; DB 2; Length 462;  
 Best Local Similarity 61.5%; Pred. No. 14;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Oy 1 DKONYALVRELK 13  
 Db 93 DSKNYLAVALQK 105

## RESULT 11

AB1088  
 chitinase B homolog lmo0105 [imported] - *Listeria monocytogenes* (strain EGD-e)

C/Species: *Listeria monocytogenes*  
 C/Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 25-Aug-2003  
 C/Accession: AB1088  
 R/Dlaer, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
 D.; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001  
 A/Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitouram, A.; Ma  
 ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,  
 A/Title: Comparative genomics of *Listeria* species.  
 A/Reference number: AB1077; MUID:21537279; PMID:11679669  
 A/Accession: AB1088  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-756 <GLA>  
 A/Cross-references: GB:NC\_003210; PIDD:CA098320.1; PID:g16409464; GSPDB:GN00177  
 A/Experimental source: strain EGD-e  
 C/Genetics:  
 A/Gene: lmo0105  
 C/Superfamily: chitinase with C-terminal chitin-binding domain

Query Match 64.1%; Score 41; DB 2; Length 756;  
 Best Local Similarity 46.2%; Pred. No. 23;  
 Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Oy 1 DKONYALVRELK 13  
 Db 198 DKONFITLQDK 210

## RESULT 12

AB1452  
 chitinase B homolog lmo0153 [imported] - *Listeria innocua* (strain Clij1262)

C/Species: *Listeria innocua*  
 C/Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 25-Aug-2003  
 C/Accession: AB1452  
 R/Glasser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
 D.; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001  
 A/Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitouram, A.; Ma  
 ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,  
 A/Title: Comparative genomics of *Listeria* species.  
 A/Reference number: AB1077; MUID:21537279; PMID:11679669  
 A/Accession: AB1452  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-756 <GLA>  
 A/Cross-references: GB:AL592022; PIDD:CA095386.1; PID:g16412572; GSPDB:GN00178

A:Experimental source: strain Clp11262

C:Genetics:

A:Gene: 110153

C:Superfamily: chitinase with C-terminal chitin-binding domain

Query Match 64.1%; Score 41; DB 2; Length 756;

Best Local Similarity 46.2%; Pred. No. 23;

Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKONTALVRELK 13

Db 198 DKONTALVRELK 210

#### RESULT 13

Probable chloroquine resistance protein CG2 (strain 7G9) - malaria parasite (Plasmodium

C:Species: Plasmodium falciparum

C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 21-Jul-2000

C:Accession: T09079

R:Su, X.Z.; Kirkman, L.A.; Fujioaka, H.; Wellems, T.E.

Cell 91, 593-603, 1997

A:Title: Complex polymorphisms in an 330 kDa protein are linked to Chloroquine-resistant

A:Reference number: Z16356; MUID:98054002; PMID:9393853

A:Accession: T09079

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2708 <STX>

A:Cross-references: EMBL:AF030692; NID:92642513; PIDN:AA047853.1; PID:92642514

A:Experimental source: strain 7G8; from Brazil

C:Genetics:

A:Gene: CG2

C:Keywords: toxin resistance

Query Match 64.1%; Score 41; DB 2; Length 2708;

Best Local Similarity 53.8%; Pred. No. 82;

Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DKONTALVRELK 13

Db 1218 DKONTALVRELK 1230

#### RESULT 14

Probable chloroquine resistance protein CG2 (strain HB3) - malaria parasite (Plasmodium

C:Species: Plasmodium falciparum

C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 21-Jul-2000

C:Accession: T09080

R:Su, X.Z.; Kirkman, L.A.; Fujioaka, H.; Wellems, T.E.

Cell 91, 593-603, 1997

A:Title: Complex polymorphisms in an 330 kDa protein are linked to Chloroquine-resistant

A:Reference number: Z16356; MUID:98054002; PMID:9393853

A:Accession: T09080

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2819 <STX>

A:Cross-references: EMBL:AF030693; NID:92642515; PIDN:AA047854.1; PID:92642516

A:Experimental source: strain HB3; from Honduras

C:Genetics:

A:Gene: CG2

C:Keywords: toxin resistance

Query Match 64.1%; Score 41; DB 2; Length 2819;

Best Local Similarity 53.8%; Pred. No. 86;

Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DKONTALVRELK 13

Db 1225 DKONTALVRELK 1237

#### RESULT 15

F7551 acyl-CoA-binding protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 19-May-2000

C:Accession: F7551

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

M.; Shen, M.; Vamathevan, J.U.; Lam, P.; McDonald, L.; Ueberback, T.; Zalewski, C.; Ma

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: F7551

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-84 <WHI>

A:Cross-references: GB:AE001879; GB:AE000513; NID:96457832; PIDN:AA09755.1; PID:9645783

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR0166

A:Map position: 1

C:Superfamily: endozepine; endozepine homology

Query Match 62.5%; Score 40; DB 2; Length 84;

Best Local Similarity 66.7%; Pred. No. 3.7;

Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 KONTALVRELK 13

Db 68 QOEYVALVRELK 79

Search completed: March 22, 2004, 07:01:26

Job time : 2.8547 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:31:13 ; Search time 0.49861 Seconds

(without alignments)  
1357.597 Million cell updates/sec

Title: US-09-662-293-7

Perfect score: 64  
Sequence: 1 DKONTALVREHK 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	46	71.9	617	1	CHIT_CABEL
2	40	62.5	383	1	CHIT_HUMAN
3	40	62.5	403	1	Y271_AQUAE
4	40	60.5	488	1	PTMB_MYCPN
5	39	60.9	509	1	YMCB_BACSV
6	38	59.4	961	1	GCSP_STRAW
7	38	59.4	961	1	GCSP_STRCO
8	38	59.4	1682	1	MSPI_PLAF3
9	38	59.4	1701	1	MSPI_PLAFW
10	38	59.4	1701	1	MSPI_PLAFW
11	37	57.8	407	1	SYR_MYCGE
12	37	57.8	731	1	UL87_HSVGA
13	37	57.8	944	1	DNIA_YEAST
14	37	57.8	964	1	PLEI_MOUSE
15	37	57.8	964	1	PLEI_MOUSE
16	37	57.8	4473	1	PLEI_CRIGR
17	37	57.8	4684	1	PLEI_HUMAN
18	36	56.2	183	1	ATPD_ANASP
19	36	56.2	505	1	DHAB_ORYSA
20	36	56.2	554	1	CHIT_MANSE
21	36	56.2	699	1	CHIT_BACCI
22	36	56.2	748	1	GUNC_PSEFL
23	36	56.2	758	1	MEYE_SYNEU
24	35	54.7	140	1	COX6_SCHPO
25	35	54.7	156	1	Y7A0_BACCR
26	35	54.7	182	1	YCX2_CVAPD
27	35	54.7	233	1	MSPI_PLAFD
28	35	54.7	290	1	BETA_YEAST
29	35	54.7	298	1	SYGA_HELPJ
30	35	54.7	303	1	SYGA_HELPJ
31	35	54.7	333	1	Y591_AQUAE
32	35	54.7	334	1	FX16_MOUSE
33	35	54.7	344	1	Y747_TREPA

34	35	54.7	361	1	HAU_MOUSE	P14433 mus musculu
35	35	54.7	369	1	RF1_SULSO	Q97966 sulfobus
36	35	54.7	399	1	SYR_MYCPN	P75122 mycoplasma
37	35	54.7	410	1	SU92_HUMAN	Q95511 homo sapien
38	35	54.7	422	1	NH54_CABEL	Q45460 caenorhabdi
39	35	54.7	444	1	NTOF_BUCAL	P57256 buchiera ap
40	35	54.7	452	1	PTB3_THEMA	Q9X066 t bifunctio
41	35	54.7	525	1	YH48_RALSO	Q8X10 ralsconla s
42	35	54.7	593	1	PA12_BOVIN	P88140 bos taurus
43	35	54.7	605	1	AOR_PYFPU	Q51739 pyrococcus
44	35	54.7	613	1	GLMS_BACPN	Q8AAB1 b glucosami
45	35	54.7	678	1	E2B6_SCHPO	P56287 schizosacch

#### ALIGNMENTS

RESULT 1  
CHIT\_CABEL STANDARD, PRT, 617 AA.  
AC Q1174; O17321;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Probable endochitinase (EC 3.2.1.14).  
GN CHT-1 OR C04F6.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Pelodieridae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Bristol N2;  
RA Nham W.;  
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 62-178 FROM N.A.  
RX MEDLINE=9832489; PubMed=9662472;  
RA de la Vega H., Specht C.A., Liu Y., Robbins P.W.;  
RT "Chitinases are a multi-gene family in Aedes, Anopheles and  
RT Drosophila."  
RL Insect Mol. Biol. 7:233-239 (1998).  
CC -I- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-  
CC acetyl-D-glucosamine polymers of chitin.  
CC -I- SIMILARITY: Belongs to chitinase class II (family 18 of glycosyl  
CC hydrolyases).  
CC -I- SIMILARITY: Contains 2 chitin-binding type-2 domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL: U42835; AAA83586.1; -  
CC EMBL: AF026152; AAB81847.1; -  
CC PIR: T15408; T15408.  
CC PIR: T37248; T37248.  
CC Wormpep: C04F6.3; C803923.  
CC InterPro: IPR002557; Chitin\_bind\_Pera.  
CC InterPro: IPR001223; Glyco\_hydro\_18.  
CC InterPro: IPR001579; Glyco\_hydro\_18AS.  
CC Pfam: PF01607; CBM\_14; 2.  
CC Pfam: PF00704; Glyco\_hydro\_18; 1.  
CC ProDom: PD000471; Glyco\_hydro\_18; 1.  
CC SMART: SM00494; ChitBD2; 2.  
CC SMART: SM00636; Glyco\_18; 1.  
CC PROSITE: PS00940; CHIT\_BIND\_II; 2.  
CC PROSITE: PS01095; CHITINASE\_18; 1.  
CC Hydrolase; Glycosidase; Chitin degradation; Chitin-binding; Repeat.  
KW DOMAIN  
FT 478 534 CHITIN-BINDING TYPE-2 1.

FT DOMAIN 563 617 CHITIN-BINDING TYPE-2.2.  
 FT ACT SITE 179 179 PROTON DONOR (BY SIMILARITY).  
 FT DOMAIN 435 478 THR-RICH.  
 FT CONFLICT 138 138 W -> C (IN REF. 2).  
 FT CONFLICT 163 165 ITP -> TTS (IN REF. 2).  
 FT CONFLICT 176 176 I -> L (IN REF. 2).  
 SQ SEQUENCE 617 AA; 66857 MW; DDAID2AACE54DA CRC64;

Query Match 71.9%; Score 46; DB 1; Length 617;  
 Best Local Similarity 69.2%; Pred. No. 0.92;  
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DKONYALVRELK 13  
 Db 186 DMANYVALRELK 198

## RESULT 2

ID C3L1\_HUMAN STANDARD; PRT; 383 AA.  
 AC P36222; P30923;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Chitinase-3 like protein 1 precursor (Cartilage glycoprotein-39) (GP-39) (39 kDa synovial protein) (YKL-40).  
 GN CH3L1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 CX NCBI\_TaxID=9606;  
 [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Cartilage;  
 RC MEDLINE=94064658; PubMed=8245017;  
 RA Hakala B.E., White C., Recklies A.D.;  
 RT "Human cartilage gp-39, a major secretory product of articular chondrocytes and synovial cells, is a mammalian member of a chitinase protein family.";  
 RL J. Biol. Chem. 268:25803-25810(1993).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Blood;  
 RC MEDLINE=97386591; PubMed=9244440;  
 RA Rehl M., Krause S.W., Andressen R.;  
 RT "Molecular characterization of the gene for human cartilage gp-39 (CH3L1), a member of the chitinase protein family and marker for late stages of macrophage differentiation.";  
 RL Genomics 43:221-225(1997).  
 [3]  
 RP SEQUENCE OF 22-45.  
 RC MEDLINE=90328983; PubMed=2375755;  
 RA Nyirkos P., Golds E.E.,  
 RT "Human synovial cells secrete a 39 kDa protein similar to a bovine mammary protein expressed during the non-lactating period.";  
 RL Biochem. J. 269:265-268(1990).  
 [4]  
 RP FUNCTION: May play an important role in the capacity of cells to respond to and cope with changes in their environment.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- TISSUE SPECIFICITY: Present in articular chondrocytes, synovial cells as well as in liver. Undetectable in muscle tissues, lung, pancreas, mononuclear cells, or fibroblasts.  
 CC -1- PTM: Glycosylated.  
 CC -1- SIMILARITY: Belongs to family 18 of glycosyl hydrolases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

DR EMBL; M80927; AA16074.1; -;  
 DR EMBL; Y08374; CA69661.1; -;  
 DR EMBL; Y08375; CA69661.1; JOINED.  
 DR EMBL; Y08376; CA69661.1; JOINED.  
 DR EMBL; Y08377; CA69661.1; JOINED.  
 DR EMBL; Y08378; CA69661.1; JOINED.  
 DR PIR; A49562; A49562.  
 DR PDB; 1LA7; 10-APR-02.  
 DR Genew; HGNC:1932; CH3L1.  
 DR MIM; 601525; -;  
 DR GO; GO:0005578; C:extracellular matrix; TAS.  
 DR GO; GO:0005615; C:extracellular space; TAS.  
 DR GO; GO:0005201; F:extracellular matrix structural constituent; TAS.  
 DR InterPro; IPR001223; Glyco\_hydro\_18.  
 DR InterPro; IPR001579; Glyco\_hydro\_18AS.  
 DR Pfam; PF00704; Glyco\_hydro\_18; 1.  
 DR PfDom; PD000471; Glyco\_hydro\_18; 1.  
 DR SMART; SM00636; Glyco\_18; 1.  
 DR PROSITE; PS01095; CHITINASE\_18; FALSE\_NEG.  
 KW Glycoprotein; Signal; 3D-structure.  
 FT SIGNAL 1 21  
 FT CHAIN 22 383 CHITINASE-3 LIKE PROTEIN 1.  
 FT CARBOHYD 60 60 N-LINKED (GLCNAC...)(POTENTIAL).  
 SQ SEQUENCE 383 AA; 42613 MW; 76AD8298BEEC2D1 CRC64;

Query Match 62.5%; Score 40; DB 1; Length 383;  
 Best Local Similarity 46.2%; Pred. No. 7.3;  
 Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DKONYALVRELK 13  
 Db 146 DKOHFTTLKEWK 158

RESULT 3  
 Y271\_AQUAE STANDARD; PRT; 403 AA.  
 AC 066628;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein AQ\_271.  
 GN AQ\_271.  
 OS Aquifex aeolicus.  
 OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.  
 CX NCBI\_TaxID=63363;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VF5;  
 RC MEDLINE=98196666; PubMed=9537320;  
 RX Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
 RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Ajay K., Huber R.,  
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
 RT "The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.";  
 RL Nature 392:353-358(1998).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

DR EMBL; AE000681; AAC06586.1; -;  
 DR PIR; H70324; H70324.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 403 AA; 46506 MW; 06C849B676F6DF01 CRC64;

Query Match 62.5%; Score 40; DB 1; Length 403;  
 Best Local Similarity 63.6%; Pred. No. 7.8;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 ONYALVIRELK 13  
 DB 107 ONYALVIRELK 117

RESULT 4  
 PTMB MYCEN STANDARD; PRT; 488 AA.

AC P75146;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE PRS system, mannitol-specific IIBC component (EIIIC-Mtl) (Mannitol-  
 permease IIBC component) (Phosphotransferase enzyme II, BC component)  
 DE (BC 2.7.1.65) (EII-Mtl).  
 GN MTLA OR MPN651 OR MP191.  
 OS Mycoplasma pneumoniae.  
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
 OX NCBI\_TaxId=2104;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 29342 / M129;  
 RX MEDLINE=97105885; PubMed=8948633;  
 RA Hammerle R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,  
 RA Hermann R.;  
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma  
 pneumoniae."  
 RT Nucleic Acids Res. 24:4420-4449(1996).  
 CC -1- FUNCTION: This is a component of the phosphoenolpyruvate-dependent  
 sugar phosphotransferase system (PRS), a major carbohydrate active  
 transport system. The IICD domains contain the sugar binding site  
 and the transmembrane channel; the IIA domain contains the primary  
 phosphorylation site (the donor is phospho-HPr); IIB transfers its  
 phosphoryl group to the IIB domain which finally transfers it to  
 the sugar (by similarity).  
 CC -1- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein  
 histidine + sugar phosphate.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (by similarity).  
 CC -1- SIMILARITY: Contains 1 PRS EIIIC domain.  
 CC -1- SIMILARITY: Contains 1 PRS EIIIC domain.  
 CC -1- CAUTION: REP-1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FRAMESHIFTS  
 IN POSITIONS 55 AND 68.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 CC EMBL, AE000020; AAB95839.1; ALT\_FRAME.  
 DR InterPro; IPR003352; Pfam\_A03501; Pfam\_A03501.  
 DR Pfam; PF02378; PTS\_EIIC.1.  
 DR Pfam; PF02302; PTS\_IIB.1.  
 KW phosphotransferase system; Sugar transport; Transferase;  
 KW Transmembrane; Phosphorylation; Complete proteome.  
 FT DOMAIN 1 352  
 FT TRANSMEM 353 488  
 FT TRANSMEM 22 42  
 FT TRANSMEM 51 71  
 FT TRANSMEM 80 100  
 FT TRANSMEM 104 124  
 FT TRANSMEM 151 171  
 FT TRANSMEM 227 247  
 FT TRANSMEM 277 297  
 FT TRANSMEM 300 320  
 FT TRANSMEM 335 355  
 FT MOD\_RES 273 273  
 FT MOD\_RES 403 403  
 FT SEQUENCE 488 AA; 53438 MW; E4CDE141110C9C4 CRC64;

QY 1 DKONYALVIRELK 13  
 DB 472 DKONYALVIRELK 484

Query Match 62.5%; Score 40; DB 1; Length 488;  
 Best Local Similarity 46.2%; Pred. No. 9.5;  
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

RESULT 5  
 YMCB BACSU STANDARD; PRT; 509 AA.

AC 031778;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical protein ymcB.  
 GN YMCB OR BSU17010.  
 OS Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxId=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunat F., Ogasawara N., Moszer I., Albertini A.M., Allioni G.,  
 RA Azevedo V., Bartorelli L., Bressan P., Bolognini S., Borcher S.,  
 RA Borris R., Boursier L., Brans A., Braun M., Briggner S.C., Bron S.,  
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Ertan K.D., Errington J., Fabre C., Ferrati E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Gallier N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Golligly E.J., Grandi G.,  
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holstappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Paro V., Pohl T.M., Portecelle D., Porwollik S., Prescott A.M.,  
 RA Plescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan B., Schleich S., Schroeder R., Scofield F.,  
 RA Sekiguchi J., Sekowska A., Serrif S.J., Serrif P., Shin B.S., Solido B.,  
 RA Sorokin A., Taccori E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Tepstra P., Tognoni A.,  
 RA Tosiato V., Uchiyama S., Vandenbol M., Vannier F., Vassart A.,  
 RA Viari A., Wambut R., Wedler E., Wedler H., Wellenreger T.,  
 RA Winfers P., Wipat A., Yamamoto H., Yamane K., Yasunaga K., Yata K.,  
 RA Yoshida K., Yoshikawa H.P., Zumbstein E., Yoshikawa H., Danchin A.,  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 subtilis."  
 RL Nature 390:249-256(1997).  
 CC -1- SIMILARITY: Belongs to the UPF0004 family.  
 CC -1- SIMILARITY: Contains 1 TRAM domain.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 CC EMBL, Z99112; CAB3574.1; -  
 DR PIR; D69884; D69884.  
 DR Subtilisin; BG13418; YMCB.  
 DR InterPro; IPR006638; EIP3.  
 DR InterPro; IPR006463; M1ab\_methiolase.

```

DR InterPro: IPR007197; Radical_SAM.
DR InterPro: IPR007192; TRAM.
DR InterPro: IPR005839; UPF0004.
DR Pfam: PF04055; Radical_SAM; 1.
DR Pfam: PF01938; TRAM; 1.
DR Pfam: PF00919; UPF0004; 1.
DR SMART: SM00729; EIP3; 1.
DR TIGRPFAMs: TIGR01574; mab-methiolase; 1.
DR TIGRPFAMs: TIGR00089; TIGR00089; 1.
DR PROSITE: PS00926; TRAM; 1.
DR PROSITE: PS01278; UPF0004; 1.
DR Hypothetical protein; Complete proteome.
KM BINDING 709 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT BINDING 709 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 509 AA; 58169 MW; 163BCE3025F196E CRC64;

Query Match 60.9%; Score 39; DB 1; Length 509;
Best Local Similarity 46.2%; Pred. No. 15;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 DKONTALVRELK 13
Db 344 DREYMEVREK 356

RESULT 6
GSP STRAM STANDARD; PRT; 961 AA.
AC 0827D7;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Glycine dehydrogenase [decarboxylating] (EC 1.4.4.2) (Glycine
decarboxylase) (Glycine cleavage system P-protein).
GN GCVB OR GCVB OR SAV6987.
OS Streptomyces avermiltis.
OC Bacteria; Actinobacteria; Actinomycetales;
OC Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa U., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hatori M.
RA "Genome sequence of an industrial microorganism Streptomyces
avermiltis: deducing the ability of producing secondary
metabolites."
RA Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Ikeda H., Ishikawa U., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hatori M., Omura S.,
RA "Complete genome sequence and comparative analysis of the industrial
microorganism Streptomyces avermiltis."
RA Nat. Biotechnol. 21:526-531(2003).
RT FUNCTION: The glycine cleavage system catalyzes the degradation of
glycine. The P protein binds the alpha-amino group of glycine
through its pyridoxal phosphate cofactor; CO(2) is released and
the remaining methylamine moiety is then transferred to the
lipamide cofactor of the H protein (By similarity).
CC -1- CATALYTIC ACTIVITY: Glycine + lipoylprotein = S-
aminoethylidihydrolypylprotein + CO(2).
CC -1- COFACTOR: Pyridoxal phosphate (By similarity).
CC -1- SUBUNIT: The glycine cleavage system is composed of four proteins:
P, T, L and H (By similarity).
CC -1- SIMILARITY: Belongs to the gcvp family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP005048; BAC74698.1; ALT_INT.
DR HAMAP: MF_00711; -; 1.
DR InterPro: IPR003437; GDC-P.
DR Pfam: PF02347; GDC-P; 1.
DR TIGRPFAMs: TIGR00461; gcvp; 1.
KM Oxidoreductase; Pyridoxal phosphate; Complete proteome.
FT BINDING 709 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 961 AA; 102649 MW; FC4CBDD2B06C32D0 CRC64;

Query Match 59.4%; Score 38; DB 1; Length 961;
Best Local Similarity 63.6%; Pred. No. 46;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 3 QNTYALVRELK 13
Db 544 QGYTLIRELE 554

RESULT 7
GSP STRCO STANDARD; PRT; 961 AA.
AC 09AK84;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Glycine dehydrogenase [decarboxylating] (EC 1.4.4.2) (Glycine
decarboxylase) (Glycine cleavage system P-protein).
GN GCVB OR SC01378 OR SC10A9.20C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinomycetales;
OC Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thompson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser H., Laikre L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wierhorrek A., Woodward J., Barrett B.G., Parkhill J.,
RA Hopwood D.A.;
RA "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2)."
RA Nature 417:141-147(2002).
RT FUNCTION: The glycine cleavage system catalyzes the degradation of
glycine. The P protein binds the alpha-amino group of glycine
through its pyridoxal phosphate cofactor; CO(2) is released and
the remaining methylamine moiety is then transferred to the
lipamide cofactor of the H protein (By similarity).
CC -1- CATALYTIC ACTIVITY: Glycine + lipoylprotein = S-
aminoethylidihydrolypylprotein + CO(2).
CC -1- COFACTOR: Pyridoxal phosphate (By similarity).
CC -1- SUBUNIT: The glycine cleavage system is composed of four proteins:
P, T, L and H (By similarity).
CC -1- SIMILARITY: Belongs to the gcvp family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

```

DR EMBL: AL939108; CAC2302.1; -  
 DR HAMAP: MF 00711; -; 1.  
 DR InterPro: IPR003437; GDC-P.  
 DR Pfam: PF02347; GDC-P; 1.  
 DR TIGRfam: TIGR00461; gcvp; 1.  
 DR Oxidoreductase; Pyridoxal phosphate; Complete proteome.  
 FT BINDING 709  
 FT SEQUENCE 961 AA; 102808 MW; 7A9D905DD1E6D35 CRC64;

Query Match 59.4%; Score 38; DB 1; Length 961;  
 Best Local Similarity 63.6%; Pred. No. 46;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 QNTALVRELK 13  
 DB 544 QGYLTIRELE 554

RESULT 8  
 MSP1\_PLAF3 STANDARD; PRT; 1682 AA.  
 AC P19598; Q25921;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens) (PMSA) (P190).  
 GN MSP-1.  
 OS Plasmodium falciparum (isolate ro-33 / Ghana).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5834;  
 RN [1]  
 RP MEDLINE=8166657; PubMed=3327688;  
 RA Certa U., Kottmann D., Matile H., Reber-Liske R.,  
 RT "A naturally occurring gene encoding the major surface antigen precursor p190 of Plasmodium falciparum lacks tripeptide repeats.";  
 RL EMBO J. 6:4137-4142(1987).  
 RN [2]  
 RP SEQUENCE OF 1032-1682 FROM N.A.  
 RX MEDLINE=9354793; PubMed=7628566;  
 RA Tolle R., Bujard H., Cooper J.A.;  
 RT "Plasmodium falciparum: Variations within the C-terminal region of merozoite surface antigen-1.";  
 RL Exp. Parasitol. 81:47-54(1995).  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (Potential).  
 CC -1- PTM: Merozoite surface antigen contain the sequence of 83 kDa, 42 kDa and 19 kDa antigens which are the major surface antigens of merozoites. The maturation take place during schizont.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M35727; AAA29715.1; -  
 CC DR EMBL: Y00087; CAA68280.1; -  
 CC DR EMBL: Z35326; CAA84555.1; -  
 CC DR InterPro: IPR006209; EGF\_Like.  
 CC DR Pfam: PF00008; EGF; 1.  
 CC KM Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;  
 CC Transmembrane; GPI-anchor.  
 FT SIGNAL 1  
 FT CHAIN 20 1682 POTENTIAL.  
 FT TRANSMEM 1662 1682 MEMBRANE ANCHOR (POTENTIAL).  
 FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 462 462 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 599 599 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 785 785 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 881 881 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 901 901 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 947 947 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1178 1178 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1569 1569 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT SEQUENCE 1682 AA; 192462 MW; C82AL1E5948CAD6 CRC64;

Query Match 59.4%; Score 38; DB 1; Length 1682;  
 Best Local Similarity 63.6%; Pred. No. 82;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 QNTALVRELK 13  
 DB 121 QNTALVRELK 131

RESULT 9  
 MSP1\_PLAF3 STANDARD; PRT; 1701 AA.  
 AC P13819;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens) (PMSA).  
 GN MSP-1.  
 OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5837;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=8142999; PubMed=2449612;  
 RA Peterson M.G., Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,  
 RA Brown G.V., Anders R.F., Kemp D.J.;  
 RT "Variation in the precursor to the major merozoite surface antigens of Plasmodium falciparum.";  
 RL Mol. Biochem. Parasitol. 27:291-302(1988).  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (Potential).  
 CC -1- PTM: Merozoite surface antigen contain the sequence of 83 kDa, 42 kDa and 19 kDa antigens which are the major surface antigens of merozoites. The maturation take place during schizont.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M19143; AAA29653.1; -  
 CC DR PIR: A54498; A54498.  
 CC DR InterPro: IPR006209; EGF\_Like.  
 CC DR Pfam: PF00008; EGF; 1.  
 CC KM Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;  
 CC Transmembrane; GPI-anchor.  
 FT SIGNAL 1  
 FT CHAIN 20 1701 POTENTIAL.  
 FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 536 536 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 607 607 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 919 919 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 991 991 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1701 AA; 193719 MW; 3920B75E73D8552 CRC64;  
Query Match 59.4%; Score 38; DB 1; Length 1701;  
Best Local Similarity 63.6%; Pred. No. 83;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 ONYALVRELK 13  
Db 127 QNYLFTIKELK 137

RESULT 10  
MSPL PLAFM STANDARD; PRT; 1701 AA.  
ID MSPL PLAFM STANDARD; PRT; 1701 AA.  
AC P08569;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Merozoite surface protein 1 precursor (Merozoite surface antigens) (PUMSA) (P190).  
GN MSP-1  
OS Plasmodium falciparum (isolate mad20 / Papua New Guinea).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
NC NCBI\_TaxID=70153;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88011243; PubMed=3079521;  
RA Tanabe K., Mackay M., Goman M., Scaife J.G.;  
RT "Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum.";  
RL J. Mol. Biol. 195:273-287(1987).  
RN [2]  
RP REVISIONS TO 1403; 1569 AND 1629.  
RA Tanabe K.;  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 1-115 FROM N.A.  
RX MEDLINE=86136024; PubMed=3004972;  
RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,  
RT Stuenkelberg H., Bujard H.;  
RL "Polymorphism of the precursor for the major surface antigens of Plasmodium falciparum merozoites: studies at the genetic level.";  
RL EMBL J. 4:3823-3829(1985).  
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (Potential).  
CC -1- PTM: Merozoite surface antigen contain the sequence of 83 kDa, 42 kDa and 19 kDa antigens which are the major surface antigens of merozoites. The maturation take place during schizont.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X05624; CAA29112.1; -  
CC InterPro; IPR006209; EGF-like.  
CC Pfam; PF00008; EGF; 1.  
CC Malaria; Merozoite; Glycoprotein; Repeat; Signal; Glycoprotein;  
CC Transmembrane; GPI-anchor.  
CC SIGNAL 1 19  
FT CHAIN 20 1701 POTENTIAL.  
FT CARBOHYD 110 110 MEROZOITE SURFACE PROTEIN 1.  
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 536 536 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 607 607 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 919 919 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 991 991 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1701 AA; 193768 MW; 3FC2E59AF96EA98 CRC64;  
Query Match 59.4%; Score 38; DB 1; Length 1701;  
Best Local Similarity 63.6%; Pred. No. 83;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 ONYALVRELK 13  
Db 127 QNYLFTIKELK 137

RESULT 11  
SYT MYCGE STANDARD; PRT; 407 AA.  
ID SYT MYCGE STANDARD; PRT; 407 AA.  
AC P47693; Q49355;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Tyrosyl-tRNA synthetase (EC 6.1.1.1) (Tyrosine--tRNA ligase) (TYRS).  
GN TYRS OR MG455.  
OS Mycoplasma genitalium.  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
NC NCBI\_TaxID=2097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=ATCC 33530 / G-37;  
RC MEDLINE=96026346; PubMed=7569993;  
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,  
RA Fleischmann R.D., Bult C.J., Kierulff A.R., Sutton G., Kelley J.M.,  
RA Fritchman J.L., Weidman J.F., Small R.V., Sandusky M., Fuhmann J.L.,  
RA Nguyen D.T., Uitterlinden T.R., Saudek D.M., Phillips C.A., Merrick J.M.,  
RA Tomb J.F., Dougherty B.A., Bock K.F., Hu P.-C., Lucier T.S.,  
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;  
RT "The minimal gene complement of Mycoplasma genitalium.";  
RL Science 270:397-403(1995).  
RN [2]  
RP SEQUENCE OF 122-202 FROM N.A.  
RX STRAIN=ATCC 33530 / G-37;  
RC MEDLINE=94075230; PubMed=8253680;  
RA Peterson S.N., Hu P.-C., Bock K.F., Hutchison C.A. III,  
RT "A survey of the Mycoplasma genitalium genome by using random sequencing.";  
RL J. Bacteriol. 175:7918-7930(1993).  
CC -1- CATALYTIC ACTIVITY: ATP + L-tyrosine + tRNA(Tyr) = AMP + dihydrophosphate + L-tyrosyl-tRNA(Tyr).  
CC -1- SUBUNIT: Homodimer (By similarity).  
CC -1- STRUCTURAL LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.  
CC -1- SIMILARITY: Contains 1 S4 RNA-binding domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U39727; AAC72475.1; -  
CC EMBL; U02247; AAA03403.1; -  
CC PIR; G64250; C64250.  
CC HSP; P00952; 2TS1.  
CC TIGR; MG455; -  
CC InterPro; IPR003942; S4.  
CC InterPro; IPR002305; tRNA-synt\_1b.  
CC InterPro; IPR001412; tRNA-synt\_1.  
CC InterPro; IPR002307; Tyr-tRNA-synt\_1b.  
CC Pfam; PF01479; S4; 1.



```

DR Pfam: P000579; tRNA-synt.1b; 1.
DR PRINTS; PRO1040; TRNASYNTHTR.
DR SMART; SM00363; S4; 1.
DR TIGRPFAM; TIGR00234; tyS; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
DR PROSITE; PS00889; S4; 1.
KW Aminoadacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 52 61 "HIGH" REGION.
FT SITE 236 240 "KMSKS" REGION.
FT BINDING 239 239 ATP (BY SIMILARITY).
FT DOMAIN 342 405 S4 RNA-BINDING.
FT CONFLICT 202 202 Q -> L (IN REF. 2).
SQ SEQUENCE 407 AA; 46830 MW; 8B0495F7DFDD738B CRC64;

Query Match 57.8%; Score 37; DB 1; Length 407;
Best Local Similarity 50.0%; Pred. No. 28;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DKONYLAVREL 12
Db 135 EKXNYEIVREL 146

RESULT 12
ID UR87 HSUSA STANDARD; PRT; 731 AA.
AC 001007;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Gene 24 protein.
GN 24.
OS Herpesvirus saimiri (strain 11).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10383;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92333688; PubMed=1321287;
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA Honess R.W.;
RT "Primary structure of the herpesvirus saimiri genome.";
RL J. Virol. 66:5047-5058(1992).
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS TOGETHER EBV BCRF1,
CC HSV-6 US8, HVS-1 24 AND HCMV UL87.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X64346; CA45647.1; -
CC InterPro; IPR004285; UL87.
CC Pfam; PF03043; UL87.1.
SQ SEQUENCE 731 AA; 83572 MW; B6B35F2A8EBBF3D3 CRC64;

Query Match 57.8%; Score 37; DB 1; Length 731;
Best Local Similarity 66.7%; Pred. No. 53;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 KONYLAVREL 13
Db 190 KONYLRLVSAK 201

RESULT 13
DNL4_YEAST STANDARD; PRT; 944 AA.
ID DNL4_YEAST

```

```

AC 008387; 002913; 002914;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA ligase II (EC 6.5.1.1) (Polydeoxyribonucleotide synthase [ATP])
DE (DNA ligase IV homolog).
GN DNL4 OR YOR005C OR UNP407 OR UNE452.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetaceae; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97051599; PubMed=8896276;
RA Sterky F., Holmberg A., Petersson B., Uhlen M.;
RT "The sequence of a 30 kb fragment on the left arm of chromosome XV
RT from Saccharomyces cerevisiae reveals 15 open reading frames, five of
RT which correspond to previously identified genes.";
RL Yeast 12:1091-1095(1996).
RN [2]
RP REVISIONS.
RA Petersson B., Sterky F., Uhlen M.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=97248602; PubMed=9092653;
RA Ramos W., Tappe N., Talamantez J., Friedberg E.C., Tomkinson A.E.;
RT "Two distinct DNA ligase activities in mitotic extracts of the yeast
RT Saccharomyces cerevisiae.";
RL Nucleic Acids Res. 25:1485-1492(1997).
CC -1- FUNCTION: HAS MINOR DNA JOINING ACTIVITY. CAN ACT ON
CC OLIGO (PDT)/POLY (RA) SUBSTRATE.
CC -1- CATALYTIC ACTIVITY: ATP + {deoxyribonucleotide} (N) +
CC {deoxyribonucleotide} (N+M) = AMP + diphosphate +
CC {SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: Belongs to the ATP-dependent DNA ligase family.
CC -1- SIMILARITY: Contains 2 BRCr domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z74913; CA99193.1; -
CC EMBL; U43491; AAC49485.1; ALT_SEQ.
CC EMBL; U43491; AAC49484.1; ALT_SEQ.
CC PIR; S66870; S66870.
CC GenOnline; 143593; -
DR SGD; S0005531; DNL4.
DR GO; GO:0003910; F:DNA ligase (ATP) activity; IDA.
DR GO; GO:006303; P:double-strand break repair via nonhomologous. .; IMP.
DR InterPro; IPR001357; BRCr.
DR InterPro; IPR000977; DNA_ligase.
DR Pfam; PF00533; BRCr; 2.
DR Pfam; PF01068; DNA_ligase; 1.
DR SMART; SM00292; BRCr; 1.
DR TIGRPFAM; TIGR00574; dnl1; 1.
DR PROSITE; PS50172; BRCr; 2.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
DR PROSITE; PS00333; DNA_LIGASE_A2; 1.
DR PROSITE; PS0160; DNA_LIGASE_A3; 1.
KW DNA repair; DNA replication; DNA recombination; Cell division; Ligase;
KW ATP-binding; Nuclear protein; Repeat.
FT DOMAIN 681 780 BRCr 1.
FT DOMAIN 840 941 BRCr 2.
FT BINDING 282 282 AMP (BY SIMILARITY).
SQ SEQUENCE 944 AA; 108514 MW; 7FB6D7927E1719B5 CRC64;

Query Match 57.8%; Score 37; DB 1; Length 944;

```

Best Local Similarity: 46.2%; Pred. No. 69;  
Matches: 6; Conservative: 5; Mismatches: 2; Indels: 0; Gaps: 0;

QY 1 DKONYALVRELK 13  
Db 216 DAQDYLSVISPDLK 228

RESULT 14  
PLE1 MOUSE STANDARD; PR1 964 AA.  
ID PLE1 MOUSE 09QX51-6; 09QX52-7; 09QX53-8; 09QX54-9; 09QX55-10; 09QX56-11; 09QX57-12; 09QX58-13; 09QX59-14; 09QX60-15; 09QX61-16; 09QX62-17; 09QX63-18; 09QX64-19; 09QX65-20; 09QX66-21; 09QX67-22; 09QX68-23; 09QX69-24; 09QX70-25; 09QX71-26; 09QX72-27; 09QX73-28; 09QX74-29; 09QX75-30; 09QX76-31; 09QX77-32; 09QX78-33; 09QX79-34; 09QX80-35; 09QX81-36; 09QX82-37; 09QX83-38; 09QX84-39; 09QX85-40; 09QX86-41; 09QX87-42; 09QX88-43; 09QX89-44; 09QX90-45; 09QX91-46; 09QX92-47; 09QX93-48; 09QX94-49; 09QX95-50; 09QX96-51; 09QX97-52; 09QX98-53; 09QX99-54; 09QX100-55; 09QX101-56; 09QX102-57; 09QX103-58; 09QX104-59; 09QX105-60; 09QX106-61; 09QX107-62; 09QX108-63; 09QX109-64; 09QX110-65; 09QX111-66; 09QX112-67; 09QX113-68; 09QX114-69; 09QX115-70; 09QX116-71; 09QX117-72; 09QX118-73; 09QX119-74; 09QX120-75; 09QX121-76; 09QX122-77; 09QX123-78; 09QX124-79; 09QX125-80; 09QX126-81; 09QX127-82; 09QX128-83; 09QX129-84; 09QX130-85; 09QX131-86; 09QX132-87; 09QX133-88; 09QX134-89; 09QX135-90; 09QX136-91; 09QX137-92; 09QX138-93; 09QX139-94; 09QX140-95; 09QX141-96; 09QX142-97; 09QX143-98; 09QX144-99; 09QX145-100; 09QX146-101; 09QX147-102; 09QX148-103; 09QX149-104; 09QX150-105; 09QX151-106; 09QX152-107; 09QX153-108; 09QX154-109; 09QX155-110; 09QX156-111; 09QX157-112; 09QX158-113; 09QX159-114; 09QX160-115; 09QX161-116; 09QX162-117; 09QX163-118; 09QX164-119; 09QX165-120; 09QX166-121; 09QX167-122; 09QX168-123; 09QX169-124; 09QX170-125; 09QX171-126; 09QX172-127; 09QX173-128; 09QX174-129; 09QX175-130; 09QX176-131; 09QX177-132; 09QX178-133; 09QX179-134; 09QX180-135; 09QX181-136; 09QX182-137; 09QX183-138; 09QX184-139; 09QX185-140; 09QX186-141; 09QX187-142; 09QX188-143; 09QX189-144; 09QX190-145; 09QX191-146; 09QX192-147; 09QX193-148; 09QX194-149; 09QX195-150; 09QX196-151; 09QX197-152; 09QX198-153; 09QX199-154; 09QX200-155; 09QX201-156; 09QX202-157; 09QX203-158; 09QX204-159; 09QX205-160; 09QX206-161; 09QX207-162; 09QX208-163; 09QX209-164; 09QX210-165; 09QX211-166; 09QX212-167; 09QX213-168; 09QX214-169; 09QX215-170; 09QX216-171; 09QX217-172; 09QX218-173; 09QX219-174; 09QX220-175; 09QX221-176; 09QX222-177; 09QX223-178; 09QX224-179; 09QX225-180; 09QX226-181; 09QX227-182; 09QX228-183; 09QX229-184; 09QX230-185; 09QX231-186; 09QX232-187; 09QX233-188; 09QX234-189; 09QX235-190; 09QX236-191; 09QX237-192; 09QX238-193; 09QX239-194; 09QX240-195; 09QX241-196; 09QX242-197; 09QX243-198; 09QX244-199; 09QX245-200; 09QX246-201; 09QX247-202; 09QX248-203; 09QX249-204; 09QX250-205; 09QX251-206; 09QX252-207; 09QX253-208; 09QX254-209; 09QX255-210; 09QX256-211; 09QX257-212; 09QX258-213; 09QX259-214; 09QX260-215; 09QX261-216; 09QX262-217; 09QX263-218; 09QX264-219; 09QX265-220; 09QX266-221; 09QX267-222; 09QX268-223; 09QX269-224; 09QX270-225; 09QX271-226; 09QX272-227; 09QX273-228; 09QX274-229; 09QX275-230; 09QX276-231; 09QX277-232; 09QX278-233; 09QX279-234; 09QX280-235; 09QX281-236; 09QX282-237; 09QX283-238; 09QX284-239; 09QX285-240; 09QX286-241; 09QX287-242; 09QX288-243; 09QX289-244; 09QX290-245; 09QX291-246; 09QX292-247; 09QX293-248; 09QX294-249; 09QX295-250; 09QX296-251; 09QX297-252; 09QX298-253; 09QX299-254; 09QX300-255; 09QX301-256; 09QX302-257; 09QX303-258; 09QX304-259; 09QX305-260; 09QX306-261; 09QX307-262; 09QX308-263; 09QX309-264; 09QX310-265; 09QX311-266; 09QX312-267; 09QX313-268; 09QX314-269; 09QX315-270; 09QX316-271; 09QX317-272; 09QX318-273; 09QX319-274; 09QX320-275; 09QX321-276; 09QX322-277; 09QX323-278; 09QX324-279; 09QX325-280; 09QX326-281; 09QX327-282; 09QX328-283; 09QX329-284; 09QX330-285; 09QX331-286; 09QX332-287; 09QX333-288; 09QX334-289; 09QX335-290; 09QX336-291; 09QX337-292; 09QX338-293; 09QX339-294; 09QX340-295; 09QX341-296; 09QX342-297; 09QX343-298; 09QX344-299; 09QX345-300; 09QX346-301; 09QX347-302; 09QX348-303; 09QX349-304; 09QX350-305; 09QX351-306; 09QX352-307; 09QX353-308; 09QX354-309; 09QX355-310; 09QX356-311; 09QX357-312; 09QX358-313; 09QX359-314; 09QX360-315; 09QX361-316; 09QX362-317; 09QX363-318; 09QX364-319; 09QX365-320; 09QX366-321; 09QX367-322; 09QX368-323; 09QX369-324; 09QX370-325; 09QX371-326; 09QX372-327; 09QX373-328; 09QX374-329; 09QX375-330; 09QX376-331; 09QX377-332; 09QX378-333; 09QX379-334; 09QX380-335; 09QX381-336; 09QX382-337; 09QX383-338; 09QX384-339; 09QX385-340; 09QX386-341; 09QX387-342; 09QX388-343; 09QX389-344; 09QX390-345; 09QX391-346; 09QX392-347; 09QX393-348; 09QX394-349; 09QX395-350; 09QX396-351; 09QX397-352; 09QX398-353; 09QX399-354; 09QX400-355; 09QX401-356; 09QX402-357; 09QX403-358; 09QX404-359; 09QX405-360; 09QX406-361; 09QX407-362; 09QX408-363; 09QX409-364; 09QX410-365; 09QX411-366; 09QX412-367; 09QX413-368; 09QX414-369; 09QX415-370; 09QX416-371; 09QX417-372; 09QX418-373; 09QX419-374; 09QX420-375; 09QX421-376; 09QX422-377; 09QX423-378; 09QX424-379; 09QX425-380; 09QX426-381; 09QX427-382; 09QX428-383; 09QX429-384; 09QX430-385; 09QX431-386; 09QX432-387; 09QX433-388; 09QX434-389; 09QX435-390; 09QX436-391; 09QX437-392; 09QX438-393; 09QX439-394; 09QX440-395; 09QX441-396; 09QX442-397; 09QX443-398; 09QX444-399; 09QX445-400; 09QX446-401; 09QX447-402; 09QX448-403; 09QX449-404; 09QX450-405; 09QX451-406; 09QX452-407; 09QX453-408; 09QX454-409; 09QX455-410; 09QX456-411; 09QX457-412; 09QX458-413; 09QX459-414; 09QX460-415; 09QX461-416; 09QX462-417; 09QX463-418; 09QX464-419; 09QX465-420; 09QX466-421; 09QX467-422; 09QX468-423; 09QX469-424; 09QX470-425; 09QX471-426; 09QX472-427; 09QX473-428; 09QX474-429; 09QX475-430; 09QX476-431; 09QX477-432; 09QX478-433; 09QX479-434; 09QX480-435; 09QX481-436; 09QX482-437; 09QX483-438; 09QX484-439; 09QX485-440; 09QX486-441; 09QX487-442; 09QX488-443; 09QX489-444; 09QX490-445; 09QX491-446; 09QX492-447; 09QX493-448; 09QX494-449; 09QX495-450; 09QX496-451; 09QX497-452; 09QX498-453; 09QX499-454; 09QX500-455; 09QX501-456; 09QX502-457; 09QX503-458; 09QX504-459; 09QX505-460; 09QX506-461; 09QX507-462; 09QX508-463; 09QX509-464; 09QX510-465; 09QX511-466; 09QX512-467; 09QX513-468; 09QX514-469; 09QX515-470; 09QX516-471; 09QX517-472; 09QX518-473; 09QX519-474; 09QX520-475; 09QX521-476; 09QX522-477; 09QX523-478; 09QX524-479; 09QX525-480; 09QX526-481; 09QX527-482; 09QX528-483; 09QX529-484; 09QX530-485; 09QX531-486; 09QX532-487; 09QX533-488; 09QX534-489; 09QX535-490; 09QX536-491; 09QX537-492; 09QX538-493; 09QX539-494; 09QX540-495; 09QX541-496; 09QX542-497; 09QX543-498; 09QX544-499; 09QX545-500; 09QX546-501; 09QX547-502; 09QX548-503; 09QX549-504; 09QX550-505; 09QX551-506; 09QX552-507; 09QX553-508; 09QX554-509; 09QX555-510; 09QX556-511; 09QX557-512; 09QX558-513; 09QX559-514; 09QX560-515; 09QX561-516; 09QX562-517; 09QX563-518; 09QX564-519; 09QX565-520; 09QX566-521; 09QX567-522; 09QX568-523; 09QX569-524; 09QX570-525; 09QX571-526; 09QX572-527; 09QX573-528; 09QX574-529; 09QX575-530; 09QX576-531; 09QX577-532; 09QX578-533; 09QX579-534; 09QX580-535; 09QX581-536; 09QX582-537; 09QX583-538; 09QX584-539; 09QX585-540; 09QX586-541; 09QX587-542; 09QX588-543; 09QX589-544; 09QX590-545; 09QX591-546; 09QX592-547; 09QX593-548; 09QX594-549; 09QX595-550; 09QX596-551; 09QX597-552; 09QX598-553; 09QX599-554; 09QX600-555; 09QX601-556; 09QX602-557; 09QX603-558; 09QX604-559; 09QX605-560; 09QX606-561; 09QX607-562; 09QX608-563; 09QX609-564; 09QX610-565; 09QX611-566; 09QX612-567; 09QX613-568; 09QX614-569; 09QX615-570; 09QX616-571; 09QX617-572; 09QX618-573; 09QX619-574; 09QX620-575; 09QX621-576; 09QX622-577; 09QX623-578; 09QX624-579; 09QX625-580; 09QX626-581; 09QX627-582; 09QX628-583; 09QX629-584; 09QX630-585; 09QX631-586; 09QX632-587; 09QX633-588; 09QX634-589; 09QX635-590; 09QX636-591; 09QX637-592; 09QX638-593; 09QX639-594; 09QX640-595; 09QX641-596; 09QX642-597; 09QX643-598; 09QX644-599; 09QX645-600; 09QX646-601; 09QX647-602; 09QX648-603; 09QX649-604; 09QX650-605; 09QX651-606; 09QX652-607; 09QX653-608; 09QX654-609; 09QX655-610; 09QX656-611; 09QX657-612; 09QX658-613; 09QX659-614; 09QX660-615; 09QX661-616; 09QX662-617; 09QX663-618; 09QX664-619; 09QX665-620; 09QX666-621; 09QX667-622; 09QX668-623; 09QX669-624; 09QX670-625; 09QX671-626; 09QX672-627; 09QX673-628; 09QX674-629; 09QX675-630; 09QX676-631; 09QX677-632; 09QX678-633; 09QX679-634; 09QX680-635; 09QX681-636; 09QX682-637; 09QX683-638; 09QX684-639; 09QX685-640; 09QX686-641; 09QX687-642; 09QX688-643; 09QX689-644; 09QX690-645; 09QX691-646; 09QX692-647; 09QX693-648; 09QX694-649; 09QX695-650; 09QX696-651; 09QX697-652; 09QX698-653; 09QX699-654; 09QX700-655; 09QX701-656; 09QX702-657; 09QX703-658; 09QX704-659; 09QX705-660; 09QX706-661; 09QX707-662; 09QX708-663; 09QX709-664; 09QX710-665; 09QX711-666; 09QX712-667; 09QX713-668; 09QX714-669; 09QX715-670; 09QX716-671; 09QX717-672; 09QX718-673; 09QX719-674; 09QX720-675; 09QX721-676; 09QX722-677; 09QX723-678; 09QX724-679; 09QX725-680; 09QX726-681; 09QX727-682; 09QX728-683; 09QX729-684; 09QX730-685; 09QX731-686; 09QX732-687; 09QX733-688; 09QX734-689; 09QX735-690; 09QX736-691; 09QX737-692; 09QX738-693; 09QX739-694; 09QX740-695; 09QX741-696; 09QX742-697; 09QX743-698; 09QX744-699; 09QX745-700; 09QX746-701; 09QX747-702; 09QX748-703; 09QX749-704; 09QX750-705; 09QX751-706; 09QX752-707; 09QX753-708; 09QX754-709; 09QX755-710; 09QX756-711; 09QX757-712; 09QX758-713; 09QX759-714; 09QX760-715; 09QX761-716; 09QX762-717; 09QX763-718; 09QX764-719; 09QX765-720; 09QX766-721; 09QX767-722; 09QX768-723; 09QX769-724; 09QX770-725; 09QX771-726; 09QX772-727; 09QX773-728; 09QX774-729; 09QX775-730; 09QX776-731; 09QX777-732; 09QX778-733; 09QX779-734; 09QX780-735; 09QX781-736; 09QX782-737; 09QX783-738; 09QX784-739; 09QX785-740; 09QX786-741; 09QX787-742; 09QX788-743; 09QX789-744; 09QX790-745; 09QX791-746; 09QX792-747; 09QX793-748; 09QX794-749; 09QX795-750; 09QX796-751; 09QX797-752; 09QX798-753; 09QX799-754; 09QX800-755; 09QX801-756; 09QX802-757; 09QX803-758; 09QX804-759; 09QX805-760; 09QX806-761; 09QX807-762; 09QX808-763; 09QX809-764; 09QX810-765; 09QX811-766; 09QX812-767; 09QX813-768; 09QX814-769; 09QX815-770; 09QX816-771; 09QX817-772; 09QX818-773; 09QX819-774; 09QX820-775; 09QX821-776; 09QX822-777; 09QX823-778; 09QX824-779; 09QX825-780; 09QX826-781; 09QX827-782; 09QX828-783; 09QX829-784; 09QX830-785; 09QX831-786; 09QX832-787; 09QX833-788; 09QX834-789; 09QX835-790; 09QX836-791; 09QX837-792; 09QX838-793; 09QX839-794; 09QX840-795; 09QX841-796; 09QX842-797; 09QX843-798; 09QX844-799; 09QX845-800; 09QX846-801; 09QX847-802; 09QX848-803; 09QX849-804; 09QX850-805; 09QX851-806; 09QX852-807; 09QX853-808; 09QX854-809; 09QX855-810; 09QX856-811; 09QX857-812; 09QX858-813; 09QX859-814; 09QX860-815; 09QX861-816; 09QX862-817; 09QX863-818; 09QX864-819; 09QX865-820; 09QX866-821; 09QX867-822; 09QX868-823; 09QX869-824; 09QX870-825; 09QX871-826; 09QX872-827; 09QX873-828; 09QX874-829; 09QX875-830; 09QX876-831; 09QX877-832; 09QX878-833; 09QX879-834; 09QX880-835; 09QX881-836; 09QX882-837; 09QX883-838; 09QX884-839; 09QX885-840; 09QX886-841; 09QX887-842; 09QX888-843; 09QX889-844; 09QX890-845; 09QX891-846; 09QX892-847; 09QX893-848; 09QX894-849; 09QX895-850; 09QX896-851; 09QX897-852; 09QX898-853; 09QX899-854; 09QX900-855; 09QX901-856; 09QX902-857; 09QX903-858; 09QX904-859; 09QX905-860; 09QX906-861; 09QX907-862; 09QX908-863; 09QX909-864; 09QX910-865; 09QX911-866; 09QX912-867; 09QX913-868; 09QX914-869; 09QX915-870; 09QX916-871; 09QX917-872; 09QX918-873; 09QX919-874; 09QX920-875; 09QX921-876; 09QX922-877; 09QX923-878; 09QX924-879; 09QX925-880; 09QX926-881; 09QX927-882; 09QX928-883; 09QX929-884; 09QX930-885; 09QX931-886; 09QX932-887; 09QX933-888; 09QX934-889; 09QX935-890; 09QX936-891; 09QX937-892; 09QX938-893; 09QX939-894; 09QX940-895; 09QX941-896; 09QX942-897; 09QX943-898; 09QX944-899; 09QX945-900; 09QX946-901; 09QX947-902; 09QX948-903; 09QX949-904; 09QX950-905; 09QX951-906; 09QX952-907; 09QX953-908; 09QX954-909; 09QX955-910; 09QX956-911; 09QX957-912; 09QX958-913; 09QX959-914; 09QX960-915; 09QX961-916; 09QX962-917; 09QX963-918; 09QX964-919; 09QX965-920; 09QX966-921; 09QX967-922; 09QX968-923; 09QX969-924; 09QX970-925; 09QX971-926; 09QX972-927; 09QX973-928; 09QX974-929; 09QX975-930; 09QX976-931; 09QX977-932; 09QX978-933; 09QX979-934; 09QX980-935; 09QX981-936; 09QX982-937; 09QX983-938; 09QX984-939; 09QX985-940; 09QX986-941; 09QX987-942; 09QX988-94



```
FT DOMAIN 1260 2544 CENTRAL FIBROUS ROD DOMAIN.
FT DOMAIN 2545 4473 GLOBULAR 2.
FT DOMAIN <1 192 ACTIN-BINDING.
FT DOMAIN <1 74 CH 1.
FT DOMAIN 87 189 CH 2.
FT REPEAT 449 508 SPECTRIN 1.
FT REPEAT 529 613 SPECTRIN 2.
FT REPEAT 626 719 SPECTRIN 3.
FT REPEAT 1104 1204 SPECTRIN 4.
FT DOMAIN 1258 2548 COILED COIL (POTENTIAL).
FT REPEAT 1258 2548 PLECTIN 1.
FT REPEAT 2615 2652 PLECTIN 2.
FT REPEAT 2653 2690 PLECTIN 3.
FT REPEAT 2691 2728 PLECTIN 4.
FT REPEAT 2729 2766 PLECTIN 5.
FT REPEAT 2770 2804 PLECTIN 6.
FT REPEAT 2805 2942 PLECTIN 7.
FT REPEAT 2943 2980 PLECTIN 8.
FT REPEAT 2981 3018 PLECTIN 9.
FT REPEAT 3019 3056 PLECTIN 10.
FT REPEAT 3057 3094 PLECTIN 11.
FT REPEAT 3274 3311 PLECTIN 12.
FT REPEAT 3312 3349 PLECTIN 13.
FT REPEAT 3350 3387 PLECTIN 14.
FT REPEAT 3388 3425 PLECTIN 15.
FT REPEAT 3429 3463 PLECTIN 16.
FT REPEAT 3609 3646 PLECTIN 17.
FT REPEAT 3647 3684 PLECTIN 18.
FT REPEAT 3685 3722 PLECTIN 19.
FT REPEAT 3723 3760 PLECTIN 20.
FT REPEAT 3764 3797 PLECTIN 21.
FT REPEAT 3800 3834 PLECTIN 22.
FT REPEAT 3852 3889 PLECTIN 23.
FT REPEAT 3890 3927 PLECTIN 24.
FT REPEAT 3928 3965 PLECTIN 25.
FT REPEAT 3966 4003 PLECTIN 26.
FT REPEAT 4007 4041 PLECTIN 27.
FT REPEAT 4043 4094 PLECTIN 28.
FT REPEAT 4197 4234 PLECTIN 29.
FT REPEAT 4235 4272 PLECTIN 30.
FT REPEAT 4273 4310 PLECTIN 31.
FT REPEAT 4311 4348 PLECTIN 32.
FT REPEAT 4349 4386 PLECTIN 33.
FT DOMAIN 4039 4089 BINDING TO INTERMEDIATE FILAMENTS (BY
FT SIMILARITY).
FT DOMAIN 4414 4429 4 X 4 AA TANDEN REPEATS OF G-S-R-X.
FT MOD RES 4328 PHOSPHORYLATION (BY CDC2).
SQ SEQUENCE 4473 AA; 509015 MW; B144615D361E3484 CRC64;
```

```
Query Match 57.8%; Score 37; DB 1; Length 4473;
Best Local Similarity 58.3%; Pred. No. 3.5e+02;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 2 KONTALVRELK 13
Db 581 KESYSALMRELE 592
```

Search completed: March 22, 2004, 06:53:01  
Job time : 1.49861 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:39:53 ; Search time 2.6159 Seconds  
(without alignment)  
1568.003 Million cell updates/sec

Title: US-09-662-293-7  
Perfect score: 64  
Sequence: 1 DKONYLALVRELK 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues  
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL.25.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mtc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeal:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64	100.0	555	5 Q9U6R7	Q9U6R7 dermatophag
2	48	75.0	1484	5 Q94298	Q94298 caenorhabdi
3	47	73.4	309	16 Q8RHG1	Q8RHG1 fusobacteri
4	47	73.4	431	5 Q8ISH5	Q8ISH5 araneus ven
5	45	70.3	176	12 Q7J923	Q7J923 adoxophyes
6	45	70.3	929	5 Q8MY79	Q8MY79 haemaphysal
7	44	68.8	435	5 Q21859	Q21859 caenorhabdi
8	43	67.2	620	5 Q9Y0D4	Q9Y0D4 penaeus mon
9	42	65.6	352	11 Q9MTV1	Q9MTV1 rattus norv
10	42	65.6	399	5 Q17840	Q17840 caenorhabdi
11	42	65.6	460	5 Q17836	Q17836 caenorhabdi
12	42	65.6	462	5 Q9W2M6	Q9W2M6 drosophila
13	42	65.6	465	5 Q17838	Q17838 caenorhabdi
14	42	65.6	522	5 Q23675	Q23675 caenorhabdi
15	42	65.6	633	5 Q22468	Q22468 caenorhabdi
16	42	65.6	831	2 Q50076	Q50076 clostridium

17	42	65.6	2078	5 Q8WPN1	Q8WPN1 cirkopleura
18	41	64.1	151	17 Q979M2	Q979M2 thermoplasm
19	41	64.1	462	16 Q55876	Q55876 synechocyst
20	41	64.1	756	16 Q92FF3	Q92FF3 listeria in
21	41	64.1	756	16 Q8VAL3	Q8VAL3 listeria mo
22	41	64.1	2708	5 Q15791	Q15791 plasmodium
23	41	64.1	2729	5 Q8IBZ6	Q8IBZ6 plasmodium
24	41	64.1	2742	5 Q15801	Q15801 plasmodium
25	41	64.1	2819	5 Q15792	Q15792 plasmodium
26	41	64.1	2838	5 Q8MP05	Q8MP05 tenebrio mo
27	40	62.5	84	16 Q8RX18	Q8RX18 tenebrio mo
28	40	62.5	383	4 Q96H17	Q96H17 homo sapien
29	40	62.5	383	4 Q81V44	Q81V44 homo sapien
30	40	62.5	460	5 Q9W2M7	Q9W2M7 drosophila
31	40	62.5	495	2 Q7X2H0	Q7X2H0 streptomyces
32	40	62.5	635	2 Q07088	Q07088 bacillus th
33	40	62.5	870	2 Q8V019	Q8V019 vibrio sp.
34	39	60.9	248	16 Q9K756	Q9K756 bacillus ha
35	39	60.9	257	5 Q7YXW0	Q7YXW0 drosophila
36	39	60.9	368	5 Q9W2M5	Q9W2M5 drosophila
37	39	60.9	441	16 Q85HR6	Q85HR6 bradyrhizob
38	39	60.9	553	5 P91731	P91731 hypantria
39	39	60.9	983	5 Q9VZV2	Q9VZV2 drosophila
40	39	60.9	1013	5 Q960M0	Q960M0 drosophila
41	38.5	60.2	342	2 Q48618	Q48618 lactococcus
42	38.5	60.2	545	16 Q98PC6	Q98PC6 rhizobium 1
43	38	59.4	46	16 Q8P082	Q8P082 streptococc
44	38	59.4	83	5 Q86RK4	Q86RK4 plasmodium
45	38	59.4	83	5 Q86R16	Q86R16 plasmodium

## ALIGNMENTS

RESULT 1	Q9U6R7	PRELIMINARY;	PRT;	555 AA.
ID	Q9U6R7	Q9U6R7		
AC	Q9U6R7	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)		
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)		
DE	98kDa HDM allergen			
OS	Dermatophagoides farinae (House-dust mite).			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;			
OC	Acariiformes; Sarcophagiformes; Astigmata; Psoroptidia; Analgoidea;			
OC	Pyroglyphidae; Dermatophagoides.			
OX	NCBI_TaxID=6554;			
RN	[1]	SEQUENCE FROM N.A.		
RP	Weber E.R., Hunter S., Steadman K., McCall C.;			
RA	"Cloning and Characterization of a 98 kDa Allergen from			
RT	Dermatophagoides farinae";			
RL	Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF187772; AAD52672.1; -			
DR	GO; GO:0005576; C:extracellular; IEA.			
DR	GO; GO:0008061; F:chitin binding; IEA.			
DR	GO; GO:0016798; F:hydrolase activity; acting on glycosyl bonds; IEA.			
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.			
DR	GO; GO:0006030; P:chitin metabolism; IEA.			
DR	InterPro; IPR002557; Chitin_bind_per.			
DR	InterPro; IPR001223; Glyco_hydro_18.			
DR	InterPro; IPR001579; Glyco_hydro_18s.			
DR	Pfam; PF00704; Glyco_hydro_18; 1_			
DR	ProDom; PD000471; Glyco_hydro_18; 1.			
DR	SMART; SM00494; GlycoD2; 1.			
DR	SMART; SM00636; Glyco_18; 1.			
DR	PROSITE; PS01095; CHITINASE_18; 1.			
KW	Glycosidase; Hydrolase.			
SQ	SEQUENCE 555 AA; 63238 MW; 0E4564A1A59B30B CRC64;			
Query Match	100.0%;	Score 64;	DB 5;	Length 555;
Best Local Similarity	100.0%;	Pred. No. 0.0025;		
Matches	13;	Conservative 0;	Mismatches 0;	Gaps 0;

QY 1 DKONYALVRELK 13  
 |||||  
 DB 172 DKONYALVRELK 184

## RESULT 2

Q94298 PRELIMINARY; PRT; 1484 AA.

AC Q94298; 01-FEB-1997 (TREMBlrel. 02, Created)  
 DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN T01C4.1  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Pelodermata; Caenorhabditis.  
 OX NCBI\_TaxId=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None.  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 investigating biology. The C. elegans Sequencing Consortium.";  
 RT Science 282:2012-2018 (1998).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RT "The sequence of C. elegans cosmid T01C4.";  
 RT Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
 RL [3]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RT "Direct Submissiion.";  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U08858; AAB09177.1; -.  
 DR FIR: T29275; T29275.  
 DR WormRep; T01C4.1; CFI2928.  
 DR GO; GO:0008061; F:chitin binding; IEA.  
 DR GO; GO:0016787; F:hydrolyase activity; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR GO; GO:0016998; P:cell wall catabolism; IEA.  
 DR InterPro; IPR001002; Chitin binding\_1.  
 DR InterPro; IPR001223; Glyco\_Hydro\_18.  
 DR InterPro; IPR001579; Glyco\_Hydro\_18AS.  
 DR InterPro; IPR002482; LysM.  
 DR Pfam; PF00704; Glyco\_hydro\_18; 1.  
 DR Pfam; PF01476; LysM; 4.  
 DR ProDom; PD000471; Glyco\_hydro\_18; 1.  
 DR SMART; SM00270; ChitinB; 2.  
 DR SMART; SM00636; Glyco\_18; 1.  
 DR SMART; SM00257; LysM; 5.  
 DR PROSITE; PS01095; CHITINASE\_18; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 1484 AA; 166596 MW; 521F69C9C485184 CRC64;

Query Match 75.0%; Score 48; DB 5; Length 1484;  
 Best Local Similarity 61.5%; Pred. No. 7.1;  
 Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKONYALVRELK 13  
 |||||  
 DB 1087 DKONYALVRELK 1099

RESULT 3  
 Q8RHG1 PRELIMINARY; PRT; 309 AA.  
 AC Q8RHG1;

DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Hypothetical protein FN2075.  
 GN FN2075.  
 OS Fusobacterium nucleatum (subsp. nucleatum).  
 OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;  
 OC Fusobacterium.  
 OX NCBI\_TaxId=76856;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 25586;  
 RX MEDLINE=2186394; PubMed=11899109;  
 RA Kaparatil V., Anderson I., Ivanova N., Reznik G., Jos T., Lykidis A.,  
 RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,  
 RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,  
 RA Larsen N., D'Souza M., Malinas T., Overbeek R.,  
 RA Fongstein M., Kyplides N., Overbeek R.;  
 RT "Genome sequence and analysis of the oral bacterium Fusobacterium  
 nucleatum strain ATCC 25586.";  
 RT J. Bacteriol. 184:2005-2018 (2002).  
 RL EMBL; AB010511; AAJ94159.1; -.  
 DR EMBL; AB010511; AAJ94159.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 309 AA; 35935 MW; 4D8866DA0D943155 CRC64;

Query Match 73.4%; Score 47; DB 16; Length 309;  
 Best Local Similarity 61.5%; Pred. No. 2.1;  
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 DKONYALVRELK 13  
 |||||  
 DB 176 DKONYALVRELK 188

## RESULT 4

Q81SH5 PRELIMINARY; PRT; 431 AA.

AC Q81SH5;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Chitinase.  
 GN Chitinase.  
 OS Araneus ventricosus.  
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;  
 OC Araneomorphae; Entelegynae; Araneidae; Araneidae; Araneus.  
 OX NCBI\_TaxId=182803;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Han J.H., Kim S.R., Sohn H.D., Jin B.R.;  
 RT "Molecular cloning of a cDNA encoding the chitinase from the spider,  
 Araneus ventricosus.";  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY120879; AAN39100.1; -.  
 DR GO; GO:0016787; F:hydrolyase activity; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR001223; Glyco\_Hydro\_18.  
 DR InterPro; IPR001579; Glyco\_Hydro\_18AS.  
 DR Pfam; PF00704; Glyco\_hydro\_18; 1.  
 DR ProDom; PD000471; Glyco\_hydro\_18; 1.  
 DR SMART; SM00636; Glyco\_18; 1.  
 DR PROSITE; PS01095; CHITINASE\_18; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 431 AA; 47238 MW; 929439397B9BC923 CRC64;

Query Match 73.4%; Score 47; DB 5; Length 431;  
 Best Local Similarity 61.5%; Pred. No. 2.9;  
 Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKONYALVRELK 13  
 |||||  
 DB 165 DKONYALVRELK 177

## RESULT 5

```

Q7923
ID 07923 PRELIMINARY; PRT; 176 AA.
AC 07923;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ORF 22.
OS Adoxophyes orana granulovirus (AoGV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae; Granulovirus.
NCBI_TaxID=170617;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22728233; PubMed=12842624;
RA Wormleaton S., Kuzio J., Winstanley D.;
RL Virology 311:350-365(2003).
DR EMBL/AF547984; AAF85659.1; -
SQ SEQUENCE 176 AA; 21065 MW; 73E2F36E0ED69440 CRC64;

Query Match 70.3%; Score 45; DB 12; Length 176;
Best Local Similarity 53.8%; Pred. No. 2.7;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKONYALVRELK 13
Db 114 DKONYSVLRNVK 126

RESULT 6
ID 08MY79 PRELIMINARY; PRT; 929 AA.
AC 08MY79;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Chitinase.
GN CHT.
OS Haemaphysalis longicornis.
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Parasitiformes; Ixodida; Ixodidae; Haemaphysalis.
NCBI_TaxID=44366;
RN [1]
RP SEQUENCE FROM N.A.
RA You M.;
RT "Molecular characterization of a chitinase protein from the hard tick
RT Haemaphysalis longicornis.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL/AB074977; BAC06447.1; -
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0008061; F:chitin binding; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0006030; P:chitin metabolism; IEA.
DR InterPro; IPR002557; Chitin bind. perz.
DR InterPro; IPR001579; Glyco_hydro_18AS.
DR Pfam; PF01607; CBM_14; 1.
DR Pfam; PF00704; Glyco_hydro_18; 2.
DR ProDom; PD000471; Glyco_hydro_18; 2.
DR SMART; SMO0494; ChCBD2; 1.
DR SMART; SMO0636; Glyco_18; 2.
DR PROSITE; PS01095; CHITINASE_18; 2.
SQ SEQUENCE 929 AA; 104423 MW; 3D70C956DC1D93C6 CRC64;

Query Match 70.3%; Score 45; DB 5; Length 929;
Best Local Similarity 61.5%; Pred. No. 16;
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKONYALVRELK 13
Db 175 DKKNVFLVRELK 187

```

```

RESULT 7
ID Q21859 PRELIMINARY; PRT; 435 AA.
AC Q21859;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE R09D1.3 protein.
GN R09D1.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews P.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z70035; CAA33863.1; -
DR PIR; T24074; T24074.
DR WormPep; R09D1.3; CE03544.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001223; Glyco_hydro_18.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR ProDom; PD000471; Glyco_hydro_18; 2.
DR SMART; SMO0636; Glyco_18; 1.
SQ SEQUENCE 435 AA; 50487 MW; 8FF171549CB1C952 CRC64;

Query Match 68.8%; Score 44; DB 5; Length 435;
Best Local Similarity 53.8%; Pred. No. 11;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DKONYALVRELK 13
Db 194 DKSNVATLREIR 206

RESULT 8
ID 09Y0D4 PRELIMINARY; PRT; 620 AA.
AC 09Y0D4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Chitinase 1.
GN CHT-1.
OS Penaeus monodon (Penaeid shrimp).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;
OC Penaeidae; Penaeus.
NCBI_TaxID=6687;
RN [1]
RP SEQUENCE FROM N.A.
RA Tan S.H., Degnan B.M., Lehnert S.A.;
RT "The Penaeus monodon chitinase I gene is differentially expressed in
RT the hepatopancreas during the moult cycle.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL/AF157503; AAD40313.1; -
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0008061; F:chitin binding; IEA.
DR GO; GO:0016798; F:hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0006030; P:chitin metabolism; IEA.
DR InterPro; IPR002557; Chitin bind. perz.
DR InterPro; IPR001223; Glyco_hydro_18.
DR InterPro; IPR001579; Glyco_hydro_18AS.

```

DR Pfam: PF00704; Glyco\_hydro\_18; 1.  
 DR Pfam: PF00704; Glyco\_hydro\_18; 1.  
 DR Pfam: PF00704; Glyco\_hydro\_18; 1.  
 DR SMART; SM00636; ChBD2; 1.  
 DR SMART; SM00636; Glyco\_18; 1.  
 DR PROSITE; PS01095; CHITINASE\_18; 1.  
 KW Glycosidase; Hydrolase; 1.  
 SQ SEQUENCE 620 AA; 69756 MW; 811170D5AC11CFE CRC64;

Query Match 67.2%; Score 43; DB 5; Length 620;  
 Best Local Similarity 61.5%; Pred. No. 24;  
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 DKONYLALVRELK 13  
 Db 178 DKONFLLVQELR 190

## RESULT 9

Q9WTV1 PRELIMINARY; PRT; 352 AA.  
 AC Q9WTV1;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Glycoprotein-39 (Fragment).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NC NCB1\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Lewis;  
 RA "Cloning of the rat homologue of Human Cartilage glycoprotein-39 a potential autoantigen in arthritis."  
 RT Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; AF062038; AAD22610.1; -  
 DR HSSP; P07254; ICTN.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR InterPro; IPR001223; Glyco\_hydro\_18.  
 DR Pfam; PF00704; Glyco\_hydro\_18; 1.  
 DR ProDom; PD000471; Glyco\_hydro\_18; 1.  
 DR SMART; SM00636; Glyco\_18; 1.  
 FT NON\_TER 1  
 FT NON\_TER 352  
 SQ SEQUENCE 352 AA; 39391 MW; CBDE991610AC936C CRC64;

Query Match 65.6%; Score 42; DB 11; Length 352;  
 Best Local Similarity 53.8%; Pred. No. 21;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 DKONYLALVRELK 13  
 Db 125 DKONFTLLKELK 137

## RESULT 10

Q17840 PRELIMINARY; PRT; 399 AA.  
 AC Q17840;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE COB89.14 protein.  
 GN COB89.14.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 NC NCB1\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Lloyd C.R.;  
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; Pubmed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for investigating biology."  
 RT Science 282:2012-2018(1998).  
 RL EMBL; Z54342; CA91153.1; -  
 DR PIR; T19115; T19115.  
 DR WormPep; COB89.14; CE02997.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR InterPro; IPR001223; Glyco\_hydro\_18.  
 DR Pfam; PF00704; Glyco\_hydro\_18; 1.  
 DR ProDom; PD000471; Glyco\_hydro\_18; 2.  
 DR SMART; SM00636; Glyco\_18; 1.  
 SQ SEQUENCE 399 AA; 45871 MW; 9E99BF2B8B2F300 CRC64;

Query Match 65.6%; Score 42; DB 5; Length 399;  
 Best Local Similarity 53.8%; Pred. No. 23;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DKONYLALVRELK 13  
 Db 163 DQNYLIFIRELR 175

RESULT 11  
 Q17836 PRELIMINARY; PRT; 460 AA.  
 AC Q17836;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE COB89.12 protein.  
 GN COB89.12.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 NC NCB1\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Lloyd C.R.;  
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; Pubmed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for investigating biology."  
 RT Science 282:2012-2018(1998).  
 RL EMBL; Z54342; CA91149.1; -  
 DR PIR; T19111; T19111.  
 DR WormPep; COB89.12; CE02997.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR InterPro; IPR001223; Glyco\_hydro\_18.  
 DR Pfam; PF00704; Glyco\_hydro\_18; 1.  
 DR ProDom; PD000471; Glyco\_hydro\_18; 2.  
 DR SMART; SM00636; Glyco\_18; 1.  
 SQ SEQUENCE 460 AA; 52809 MW; 6984C059D47E3B69 CRC64;

Query Match 65.6%; Score 42; DB 5; Length 460;  
 Best Local Similarity 53.8%; Pred. No. 27;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DKONYLALVRELK 13  
 Db 229 DENNYLMFIRELR 241



RESULT 12  
Q9W2M6 PRELIMINARY; PRT; 462 AA.  
ID Q9W2M6 Q9W2M6; (TREMblrel. 13, Created)  
AC Q9W2M6 Q9W2M6; (TREMblrel. 22, Last sequence update)  
DT 01-OCT-2002 (TREMblrel. 25, Last annotation update)  
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)  
DE CG3986 protein (R662779p).  
GN CRT4 OR CG3986.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RX MEDLINE=20196006; PubMed=10731132;  
RA STRAIN=Berkley;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amaratunga C., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brannon R.C., Rogers Y.-H.C., Blasei R.G., Champe M., Pfeiffer B.D.,  
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.W., Basu A., Baxendale J., Bayraktaroglu L., Beasley S.K.,  
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhattacharya D., Bolintinas S.,  
RA Botvina D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,  
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Dey A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jajalil M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodra C.D., Kraft C., Kravitz S., Kulp D., Lai X.,  
RA Laetzo P., Lei Y., Levytsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei E., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Svirskas R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of Drosophila melanogaster.";  
RU Science 287:2185-2195(2000).  
[12]  
RP SEQUENCE FROM N.A.  
RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Amaratunga C., Scherer S.E.,  
RA Barton J., An H., Baldwin D., Barton J., Beeson K.Y., Busan D.A.,  
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dreanek D., Farfan D.,  
RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jajalil M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Patel S., Pfeiffer B.,  
RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,  
RA Phoonanavong S., Piltman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svirskas R., Tecor C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,  
RT "Sequencing of Drosophila melanogaster genome.";  
RU Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.

RP [3]  
RA SEQUENCE FROM N.A.  
RA Maira S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminer J.S., Prochuk S.E., Smith C.D.,  
RA Tupy J.L., Bergman C., Bernan B., Carlson J.W., Celniker S.E.,  
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,  
RT Annotation of Drosophila melanogaster genome.  
RU Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
RP [4]  
RA SEQUENCE FROM N.A.  
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.,  
RU Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
RP [5]  
RA SEQUENCE FROM N.A.  
RA Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.  
RP [6]  
RA SEQUENCE FROM N.A.  
RA STRAIN=Berkley;  
RA Chapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Dreanek D., Farfan D., Frise E.,  
RA George R., Gonzalez M., Guarni H., Krommiller B., Li P., Liao G.,  
RA Miranda A., Mungall C.J., Nunoo J., Pacle J., Paragas V., Park S.,  
RA Patel S., Phoonanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
RA Celniker S.,  
RU Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.  
RA EMBL: A803452; AAF4664.2; -;  
RA FLYBASE: AY071859; AAL49181.1; -;  
RA FLYBASE: FBgn0022700; Cht4.  
DR GO: GO:0005576; C:extracellular; IEA.  
DR GO: GO:0008061; F:chitin binding; IEA.  
DR GO: GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.  
DR GO: GO:0005875; P:carbohydrate metabolism; IEA.  
DR GO: GO:0006030; P:chitin metabolism; IEA.  
DR InterPro: IPR001257; Chitin bind Pept.  
DR InterPro: IPR001223; Glyco\_hydro\_18.  
DR InterPro: IPR001579; Glyco\_hydro\_18AS.  
DR Pfam: PF01607; CEM\_14; 1.  
DR Pfam: PF00704; Glyco\_hydro\_18; 1.  
DR ProDom: PD000471; Glyco\_hydro\_18; 1.  
DR SMART: SMO0494; ChtBD2; 1.  
DR PROSITE: PS01095; CHITINASE\_18; 1.  
KM Glycosidase; Hydrolase.  
SQ SEQUENCE 462 AA; 49858 MW; 3F9AA7D59536D1F1 CRC64;  
Query Match 65.6%; Score 42; DB 5; Length 462;  
Best Local Similarity 46.2%; Pred. No. 27;  
Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;  
QY 1 DKONYALVREIK 13  
Db 159 DRENFTLLREIK 171  
[12]  
RP SEQUENCE FROM N.A.  
RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Amaratunga C., Scherer S.E.,  
RA Barton J., An H., Baldwin D., Barton J., Beeson K.Y., Busan D.A.,  
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dreanek D., Farfan D.,  
RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jajalil M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Patel S., Pfeiffer B.,  
RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,  
RA Phoonanavong S., Piltman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svirskas R., Tecor C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,  
RT "Sequencing of Drosophila melanogaster genome.";  
RU Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.

RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
 RN ID Q22468 PRELIMINARY; PRT; 633 AA.  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none.  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology.";  
 RL Science 282:2012-2018(1998).  
 DR EMBL; Z54342; CAA91151.1; -.  
 DR PIR; T19113; T19113.  
 DR WormRep; CO8H9.4; CE02990.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR InterPro; IPR001223; Glyco\_hydro\_18.  
 DR Pfam; PF00704; Glyco\_hydro\_18; 1.  
 DR ProDom; PD000471; Glyco\_hydro\_18; 2.  
 DR SMART; SM00636; Glyco\_18; 1.  
 SQ SEQUENCE 465 AA; 53325 MW; 6BC4DC697184F23C CRC64;

Query Match 65.6%; Score 42; DB 5; Length 465;  
 Best Local Similarity 53.8%; Pred. No. 28;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DKONYLALVRELK 13  
 |||||:  
 Db 234 DKNYLMFIRELR 246

RESULT 14  
 ID Q23675 PRELIMINARY; PRT; 522 AA.  
 AC Q23675;  
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)  
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)  
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)  
 DE ZK938.6 protein.  
 GN ZK938.6  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RL lloyd C.R.;  
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
 RP Submited (JUN-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RX SEQUENCE FROM N.A.  
 RM MEDLINE=99069613; PubMed=9851916;  
 RA none.  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology.";  
 RL Science 282:2012-2018(1998).  
 DR EMBL; Z49913; CAA90145.1; -.  
 DR PIR; T28113; T28113.  
 DR WormRep; ZK938.6; CE02400.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR InterPro; IPR001223; Glyco\_hydro\_18.  
 DR Pfam; PF00704; Glyco\_hydro\_18; 2.  
 DR ProDom; PD000471; Glyco\_hydro\_18; 3.  
 DR SMART; SM00636; Glyco\_18; 1.  
 SQ SEQUENCE 522 AA; 60207 MW; 38646B4B59323215 CRC64;

Query Match 65.6%; Score 42; DB 5; Length 522;  
 Best Local Similarity 53.8%; Pred. No. 31;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DKONYLALVRELK 13  
 |||||:  
 Db 291 DENNYLMFIRELR 303

RESULT 15

Q22468  
 ID Q22468 PRELIMINARY; PRT; 633 AA.  
 AC Q22468;  
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)  
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
 DE T19113.3 protein.  
 GN T19113.3  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RX SEQUENCE FROM N.A.  
 RM MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology.";  
 RL Science 282:2012-2018(1998).  
 DR EMBL; Z66524; CAA91419.1; -.  
 DR PIR; T24898; T24898.  
 DR HSSP; P27275; IMWC.  
 DR WormRep; T19113.3; CE03654.  
 DR GO; GO:0008061; F:chitin binding; IEA.  
 DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR001002; Chitin binding\_1.  
 DR InterPro; IPR001223; Glyco\_hydro\_18.  
 DR InterPro; IPR001579; Glyco\_hydro\_18AS.  
 DR Pfam; PF00704; Glyco\_hydro\_18; 1.  
 DR ProDom; PD000471; Glyco\_hydro\_18; 1.  
 DR SMART; SM00270; ChcBD1; 3.  
 DR SMART; SM00636; Glyco\_18; 1.  
 DR PROSITE; PS01095; CHITINASE\_18; 1.  
 DR GlycoSIDase; Hydrolase.  
 KM  
 SQ SEQUENCE 633 AA; 70746 MW; 5D86F7E12FCD62C5 CRC64;

Query Match 65.6%; Score 42; DB 5; Length 633;  
 Best Local Similarity 53.8%; Pred. No. 38;  
 Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 DKONYLALVRELK 13  
 |||||:  
 Db 164 DRANYVLMRELK 176  
 Search completed: March 22, 2004, 06:59:20  
 Job time : 4.6159 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:30:23 ; Search time 3.95275 Seconds  
(without alignments)  
929.256 Million cell updates/sec

Title: US-09-662-293-7  
Perfect score: 64  
Sequence: 1 DKONTALVRLK 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

1:	Geneseqp1980s:*
2:	Geneseqp1990s:*
3:	Geneseqp2000s:*
4:	Geneseqp2001s:*
5:	Geneseqp2002s:*
6:	Geneseqp2003as:*
7:	Geneseqp2003bs:*
8:	Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64	100.0	13	AAV52516	House dus
2	64	100.0	13	AAU96320	Der HMW-m
3	64	100.0	536	AAV52525	House dus
4	64	100.0	536	AAU96329	Der HMW-m
5	64	100.0	555	AAV52523	House dus
6	64	100.0	555	AAU96327	Der HMW-m
7	64	100.0	555	AAU96328	Der HMW-m
8	60	93.8	490	AAV52535	D. pteron
9	60	93.8	490	AAU96339	Der HMW-m
10	60	93.8	509	AAV52533	D. pteron
11	60	93.8	509	AAU96337	Der HMW-m
12	60	93.8	509	AAU96338	Der HMW-m
13	45	70.3	929	ADB78972	Tick chit
14	41	64.1	756	ABB49221	Listeria
15	41	64.1	1042	ABG1702	Novel hum
16	40	62.5	258	ABG13045	Novel hum
17	40	62.5	281	ABG03900	Novel hum
18	40	62.5	281	ABG13960	Novel hum
19	40	62.5	305	ADCI14212	Human enz
20	40	62.5	308	ABR41620	Human DIT
21	40	62.5	321	ABR81342	Polypepti
22	40	62.5	327	ADCI14208	Human ova
23	40	62.5	383	ABG96297	Human enz
24	40	62.5	383	ABU56651	Lung canc
25	40	62.5	383	ABU89725	Protein d

26	40	62.5	460	4	ABB64366	Abd64366 Drosophi1
27	39	60.9	115	4	ABG05109	Abg05109 Novel hum
28	39	60.9	115	4	ABG14423	Abg14423 Novel hum
29	39	60.9	115	4	ABG14887	Abg14887 Novel hum
30	39	60.9	115	4	ABG03849	Abg03849 Novel hum
31	39	60.9	115	4	ABG05403	Abg05403 Novel hum
32	39	60.9	289	3	AAV68247	Murine cl
33	39	60.9	289	3	AAV52901	Murine cl
34	39	60.9	289	4	AAV58662	Murine cl
35	39	60.9	368	4	ABG65118	Drosophi1
36	39	60.9	480	4	ABG06842	Novel hum
37	39	60.9	534	4	ABG05413	Novel hum
38	39	60.9	548	4	ABG05422	Novel hum
39	39	60.9	553	6	ABP72626	Hyphanti1
40	39	60.9	696	4	ABR58615	AbR58615 Drosophi1
41	39	60.9	770	4	ABG12547	Novel hum
42	39	60.9	860	4	ABG08690	Novel hum
43	39	60.9	1015	4	AAU32370	Novel hum
44	39	60.9	1015	4	AAU30929	Novel hum
45	39	60.9	1102	4	ABG27548	Novel hum

ALIGNMENTS

RESULT 1  
AAV52516  
ID AAV52516 standard; peptide; 13 AA.  
AC AAV52516;  
DT 22-FEB-2000 (first entry)  
XX  
DE House dust mite allergen protein (map) A/B fragment map(6).  
XX  
KM Mite allergen protein; map; high molecular weight; HMW-map; allergy;  
KM house dust mite; IGE; immunoglobulin E; allergen; mapA; mapB;  
KM hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
KM canine; veterinary; antibody; vaccine; immunisation.  
XX  
OS Dermatophagoides farinae.  
XX  
PN MO9954349-A2.  
PD 28-OCT-1999.  
PF 16-APR-1999; 99MO-US008524.  
XX  
PR 17-APR-1998; 98US-0062013.  
PR 13-MAY-1998; 98US-0085295P.  
PR 02-SEP-1998; 98US-0098909P.  
XX  
PA (HESK-) HESKA CORP.  
XX  
PI Mccall CA, Hunter SW, Weber ER;  
XX  
DR WPI, 2000-052700/04.  
XX  
PT Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
PT used to modify an animals' hypersensitivity to mite allergens.  
XX  
PS Claim 3, Page 69; 154pp; English.  
XX  
CC Sequences AAV52510-Y52522 represent proteolytic fragments of  
CC Dermatophagoides farinae high molecular weight mite allergen protein (HMW  
CC map) composition. The HMW-map composition was isolated from a D. farinae  
CC homogenate by gel filtration, with each fraction being analysed for the  
CC presence of proteins that bound to Ige present in mite-allergic dog  
CC antisera. The HMW-map composition comprises mapA (a 109 kD protein) and  
CC mapB (98 kD). Mite allergenic proteins and peptides, and nucleic acids  
CC encoding them, may be used in therapeutic compositions to modify an  
CC animal's hypersensitivity reaction to mite allergens. Animals that may be  
CC treated include mammals and birds, especially felines, canines, equines.

CC humans, other pets, and work or domestic animals. The proteins or  
 CC fragments may also be used to diagnose allergies via a skin test. The  
 CC proteins and peptides can also be used to raise antibodies, which have a  
 CC variety of potential uses. For example, they can be used as vaccines to  
 CC passively immunise animals against dust mite hypersensitivity, as  
 CC positive controls in test kits and as tools to recover desired dust mite  
 CC allergens from a mixture of proteins

XX SQ Sequence 13 AA;

Query Match 100.0%; Score 64; DB 3; Length 13;

Best Local Similarity 100.0%; Pred. No. 8.1e-05; Mismatches 0; Gaps 0;

QY 1 DKONTALVRELK 13  
 DB 1 DKONTALVRELK 13

RESULT 2  
 AAU96320  
 ID AAU96320 standard; peptide; 13 AA.

XX AAU96320;

XX 15-JUL-2002 (first entry)

XX Der HMW-map polypeptide #7.

XX Der HMW-map; American house dust mite; antiallergic; mite; IGE;

XX mite allergenic protein; immunoglobulin E; hypersensitivity;

XX immunocomplex formation.

XX Dermatophagoides farinae.

XX WO200222807-A2.

XX 21-MAR-2002.

XX 14-SEP-2001; 2001WO-US028730.

XX 14-SEP-2000; 2000US-00662293.

XX (HESK-) HESKA CORP.

XX Mccall CA, Hunter SW, Weber ER;

XX WPI; 2002-351888/38.

XX New mite allergenic protein isolated from Dermatophagoides, designated

XX Der HMW-map protein, useful as a vaccine for treating mite allergy.

XX Claim 12; Page 70; 161pp; English.

XX The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic  
 CC acid. The Der HMW-map protein is useful for eliciting an immune response  
 CC against Der HMW-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitising a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig)E or Der HMW-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HMW-map polypeptides of the invention

XX Sequence 13 AA;

Query Match 100.0%; Score 64; DB 5; Length 13;

Best Local Similarity 100.0%; Pred. No. 8.1e-05; Mismatches 0; Gaps 0;

QY 1 DKONTALVRELK 13  
 DB 1 DKONTALVRELK 13

RESULT 3  
 AA52525

ID AA52525 standard; protein; 536 AA.

XX AA52525;

XX 22-FEB-2000 (first entry)

XX House dust mite (D. farinae) mite allergen protein (map) Pderf98-536.

XX Mite allergen protein; map; high molecular weight; HMW-map; allergy;

XX house dust mite; IGE; immunoglobulin E; allergen; mapA; mapB;

XX hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;

XX canine; veterinary; antibody; vaccine; immunisation.

XX Dermatophagoides farinae.

XX WO9954349-A2.

XX 16-APR-1999; 99WO-US008524.

XX 17-APR-1998; 98US-00062013.

XX 13-MAY-1998; 98US-0085235P.

XX 02-SEP-1998; 98US-0098909P.

XX (HESK-) HESKA CORP.

XX Mccall CA, Hunter SW, Weber ER;

XX WPI; 2000-052700/04.

XX N-PEDB; AA238579; AA238580.

XX Novel high molecular weight Dermatophagoides nucleic acid polypeptides

XX used to modify an animals' hypersensitivity to mite allergens.

XX Claim 3; Page 125-127; 154pp; English.

XX This sequence represents Dermatophagoides farinae mite allergen protein  
 CC (map) Pderf98-536, the mature form of Pderf98-555 (AA52523). Pderf98-536  
 CC has a molecular weight of 98 kD, comprising 536 amino acids, and is a  
 CC component of the Dermatophagoides farinae high molecular weight mite  
 CC allergen protein (HMW-map) composition. The HMW-map composition was  
 CC isolated from a D. farinae homogenate by gel filtration, with each  
 CC fraction being analysed for the presence of proteins that bound to IgE  
 CC present in mite-allergic dog antisera. Mite allergenic proteins and  
 CC peptides, and nucleic acids encoding them, may be used in therapeutic  
 CC compositions to modify an animal's hypersensitivity reaction to mite  
 CC allergens. Animals that may be treated include mammals and birds,  
 CC especially felines, canines, equines, humans, other pets, and work or  
 CC domestic animals. The proteins or fragments may also be used to diagnose  
 CC allergies via a skin test. The proteins and peptides can also be used to  
 CC raise antibodies, which have a variety of potential uses. For example,  
 CC they can be used as vaccines to passively immunise animals against dust  
 CC mite hypersensitivity, as positive controls in test kits and as tools to  
 CC recover desired dust mite allergens from a mixture of proteins

XX Sequence 536 AA;

Query Match 100.0%; Score 64; DB 3; Length 536;

Best Local Similarity 100.0%; Pred. No. 0.0057; Mismatches 0; Gaps 0;

QY 1 DKONTALVRELK 13



AC AAU96327;  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX  
 DE Der HMW-map polypeptide #14.  
 XX  
 KM Der HMW-map; American house dust mite; anti-allergic; mite; IgE;  
 KM mite allergenic protein; immunoglobulin E; hypersensitivity;  
 XX immunocomplex formation.  
 OS Dermatophagoides farinae.  
 XX  
 PN WO200222807-A2.  
 XX  
 PD 21-MAR-2002.  
 XX  
 PF 14-SEP-2001; 2001WO-US028730.  
 XX  
 PR 14-SEP-2000; 2000US-00662293.  
 XX  
 PA (HESK-) HESKA CORP.  
 XX  
 PI Mccall CA, Hunter SW, Weber ER;  
 XX  
 DR WPI: 2002-351888/38.  
 DR N-PSDB; ABK69571.  
 XX  
 PT New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HMW-map protein, useful as a vaccine for treating mite allergy.  
 XX  
 PS Claim 12; Page 114-116; 161pp; English.  
 XX  
 CC The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic  
 CC acid. The Der HMW-map protein is useful for eliciting an immune response  
 CC against Der HMW-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitizing a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig)E or Der HMW-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HMW-map polypeptides of the invention  
 XX  
 SQ Sequence 555 AA;

Query Match 100.0%; Score 64; DB 5; Length 555;  
 Best Local Similarity 100.0%; Pred. No. 0.0059;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 DKONTALVRELK 13  
 |||||  
 DB 172 DKONTALVRELK 184

RESULT 7

AAU96328  
 ID AAU96328 standard; protein; 555 AA.

AC AAU96328;

DT 15-JUL-2002 (first entry)

DE Der HMW-map polypeptide #15.

KM Der HMW-map; American house dust mite; anti-allergic; mite; IgE;  
 KM mite allergenic protein; immunoglobulin E; hypersensitivity;  
 XX immunocomplex formation.

OS Dermatophagoides farinae.  
 XX  
 PN WO200222807-A2.  
 XX  
 PD 21-MAR-2002.  
 XX  
 PF 14-SEP-2001; 2001WO-US028730.  
 XX  
 PR 14-SEP-2000; 2000US-00662293.  
 XX  
 PA (HESK-) HESKA CORP.  
 XX  
 PI Mccall CA, Hunter SW, Weber ER;  
 XX  
 DR WPI: 2002-351888/38.  
 DR N-PSDB; ABK69573.  
 XX  
 PT New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HMW-map protein, useful as a vaccine for treating mite allergy.  
 XX  
 PS Claim 12; Page 120-122; 161pp; English.  
 XX  
 CC The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic  
 CC acid. The Der HMW-map protein is useful for eliciting an immune response  
 CC against Der HMW-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitizing a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig)E or Der HMW-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HMW-map polypeptides of the invention  
 XX  
 SQ Sequence 555 AA;

Query Match 100.0%; Score 64; DB 5; Length 555;  
 Best Local Similarity 100.0%; Pred. No. 0.0059;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 DKONTALVRELK 13  
 |||||  
 DB 172 DKONTALVRELK 184

RESULT 8

AAU52535  
 ID AAU52535 standard; protein; 490 AA.

AC AAU52535;

DT 06-AUG-2003 (revised)

DT 22-FEB-2000 (first entry)

DE D. pteronyssinus 98 kD mite allergen protein (map) Pderp98-490.

KM Mite allergen protein; map; high molecular weight; HMW-map; allergy;  
 KM house dust mite; IgE; immunoglobulin E; allergen; mapB;

KM hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
 KM canine; veterinary; antibody; vaccine; immunisation.

OS Dermatophagoides pteronyssinus.

XX  
 XX Key Location/Qualifiers

FT Modified-site 115..117  
 FT Modified-site /note="Asn is N-glycosylated"  
 FT Modified-site 240..242  
 FT /note="Asn is N-glycosylated"

```

XX MN0954349-A2.
XX 28-OCT-1999.
XX PF 16-APR-1999; 99MO-US008524.
XX 17-APR-1998; 98US-00062013.
XX 13-MAY-1998; 98US-0085295P.
XX 02-SEP-1998; 98US-0098909P.
XX (HESK-) HESKA CORP.
XX Mccall CA, Hunter SW, Weber ER;
XX WPI; 2000-052700/04.
XX N-PSDB; AAZ38589, AAZ38590.
XX Novel high molecular weight Dermatophagoides nucleic acid polypeptides
XX used to modify an animals' hypersensitivity to mite allergens.
XX Claim 3; Page 147-149; 154pp; English.
XX This sequence represents Dermatophagoides pteronyssinus mite allergen
XX protein (map) Pderp98-490, the mature form of pderp98-509. Pderp98-490
XX has a molecular weight of 98 kD, comprising 490 amino acids, and has a
XX high degree of homology with the D. farinae mature 98 kD allergen, map3
XX (AAV52525). Nucleic acid molecules encoding Pderp98-490 were isolated
XX from a D. pteronyssinus cDNA library by hybridization with a probe
XX encoding the D. farinae high molecular weight map (HMW-map) composition.
XX Mite allergenic proteins and peptides, and nucleic acids encoding them,
XX may be used in therapeutic compositions to modify an animal's
XX hypersensitivity reaction to mite allergens. Animals that may be treated
XX include mammals and birds, especially felines, canines, equines, humans,
XX other pets, and work or domestic animals. The proteins or fragments may
XX also be used to diagnose allergies via a skin test. The proteins and
XX peptides can also be used to raise antibodies, which have a variety of
XX potential uses. For example, they can be used as vaccines to passively
XX immunize animals against dust mite hypersensitivity, as positive controls
XX in test kits and as tools to recover desired dust mite allergens from a
XX mixture of proteins. (Updated on 06-AUG-2003 to correct OS field.)
XX
SQ Sequence 490 AA;
Query Match 93.8%; Score 60; DB 3; Length 490;
Best Local Similarity 92.3%; Pred. No. 0.028;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0
CY 1 DKONTLVRELK 13
ID ||||| |||||
ID AAU96339
ID AAU96339 standard; protein; 490 AA.
XX AAU96339;
XX 15-JUL-2002 (first entry)
XX Der HMW-map polypeptide #26.
XX Der HMW-map; American house dust mite; anti-allergic; mite; IGB;
XX mite allergenic protein; immunoglobulin E; hypersensitivity;
XX immunocomplex formation.
XX Dermatophagoides farinae.
XX WO200222807-A2.
XX 21-MAR-2002.
XX 14-SEP-2001; 2001WO-US028730.
XX

```

[illegible]

PR 13-MAY-1998; 98US-0085295P.  
 PR 02-SEP-1998; 98US-0098909P.  
 PA (HESK-) HESKA CORP.  
 PI McCall CA, Hunter SW, Weber ER;  
 XX  
 XX  
 DR WPI: 2000-052700/04.  
 DR N-PSDB; AA238585, AA238586, AA238587, AA238588.  
 XX  
 PT Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
 used to modify an animals' hypersensitivity to mite allergens.  
 XX  
 PS Claim 3; Page 134-136; 154pp; English.  
 XX  
 CC This sequence represents Dermatophagoides pteronyssinus mite allergen  
 CC protein (map) Pderp98-509. Pderp98-509 has a molecular weight of 98 kD,  
 CC comprising 509 amino acids, and has a high degree of homology with the D.  
 CC farinae 98 kD allergen, mapB (AA252523). Nucleic acid molecules encoding  
 CC Pderp98-509 were isolated from a D. pteronyssinus cDNA library by  
 CC hybridisation with a probe encoding the D. farinae high molecular weight  
 CC map (HWM-map) composition. Mite allergenic proteins and peptides, and  
 CC nucleic acids encoding them, may be used in therapeutic compositions to  
 CC modify an animal's hypersensitivity reaction to mite allergens. Animals  
 CC that may be treated include mammals and birds, especially felines, the  
 CC canines, equines, humans, other pets, and work or domestic animals. The  
 CC proteins or fragments may also be used to diagnose allergies via a skin  
 CC test. The proteins and peptides can also be used to raise antibodies,  
 CC which have a variety of potential uses. For example, they can be used as  
 CC vaccines to passively immunise animals against dust mite  
 CC hypersensitivity, as positive controls in test kits and as tools to  
 CC recover desired dust mite allergens from a mixture of proteins. (Updated  
 CC on 06-AUG-2003 to correct OS field.)  
 CC  
 SQ Sequence 509 AA;

Query Match 93.8%; Score 60; DB 3; Length 509;  
 Best Local Similarity 92.3%; Pred. No. 0.029;  
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKONYTALVRELK 13  
 |||||  
 DB 172 DKONYTALVRELK 184

## RESULT 11

AAU96337  
 ID AAU96337 standard; protein; 509 AA.

AC AAU96337;  
 XX

DT 15-JUL-2002 (first entry)  
 XX

DE Der HWM-map polypeptide #24.  
 XX

KM Der HWM-map; American house dust mite; antiallergic; mite; IGE;  
 KM mite allergenic protein; immunoglobulin E; hypersensitivity;  
 KM immunocomplex formation.  
 XX

OS Dermatophagoides farinae.  
 XX

PN WO200222807-A2.  
 XX

PD 21-MAR-2002.  
 XX

PF 14-SEP-2001; 2001WO-US028730.  
 XX

PR 14-SEP-2000; 2000US-00662293.  
 XX

PA (HESK-) HESKA CORP.  
 XX

PI McCall CA, Hunter SW, Weber ER;  
 XX

DR WPI: 2002-351888/38.  
 DR N-PSDB; ABK69581.  
 XX  
 PT New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HWM-map protein, useful as a vaccine for treating mite allergy.  
 XX  
 PS Claim 12; Page 134-136; 161pp; English.  
 XX  
 CC The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HWM-map protein, and its related nucleic  
 CC acid. The Der HWM-map protein is useful for eliciting an immune response  
 CC against Der HWM-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitising a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HWM-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig)E or Der HWM-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HWM-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96334-AAU96342  
 CC represent Der HWM-map polypeptides of the invention  
 CC  
 SQ Sequence 509 AA;

Query Match 93.8%; Score 60; DB 5; Length 509;  
 Best Local Similarity 92.3%; Pred. No. 0.029;  
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKONYTALVRELK 13  
 |||||  
 DB 172 DKONYTALVRELK 184

## RESULT 12

AAU96338  
 ID AAU96338 standard; protein; 509 AA.

AC AAU96338;  
 XX

DT 15-JUL-2002 (first entry)  
 XX

DE Der HWM-map polypeptide #25.  
 XX

KM Der HWM-map; American house dust mite; antiallergic; mite; IGE;  
 KM mite allergenic protein; immunoglobulin E; hypersensitivity;  
 KM immunocomplex formation.  
 XX

OS Dermatophagoides farinae.  
 XX

PN WO200222807-A2.  
 XX

PD 21-MAR-2002.  
 XX

PF 14-SEP-2001; 2001WO-US028730.  
 XX

PR 14-SEP-2000; 2000US-00662293.  
 XX

PA (HESK-) HESKA CORP.  
 XX

PI McCall CA, Hunter SW, Weber ER;  
 XX

DR WPI: 2002-351888/38.  
 XX

DR N-PSDB; ABK69583.  
 XX

PT New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HWM-map protein, useful as a vaccine for treating mite allergy.  
 XX

PS Claim 12; Page 139-141; 161pp; English.  
 XX

CC The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HWM-map protein, and its related nucleic



CC acid. The Der HMM-map protein is useful for eliciting an immune response  
 CC against Der HMM-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitizing a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMM-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig)E or Der HMM-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HMM-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HMM-map polypeptides of the invention

XX SQ Sequence 509 AA;

Query Match 93.8%; Score 60; DB 5; Length 509;  
 Best Local Similarity 92.3%; Pred. No. 0.029;  
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKONYLALVRELK 13  
 Db 172 DKONYLTLVRELK 184

#### RESULT 13

ID ADB78972 standard; protein; 929 AA.

XX ADB78972;

DT 04-DEC-2003 (first entry)

DE Tick chitinase #SEQ ID 2.

XX Chitinase; antiinflammatory; virucide; antipyretic; tick; mite;

KM piroplasmosis; Q fever; viral encephalitis; vaccine.

XX Haemaphysalis longicornis.

XX MO2003072609-AA.

XX 04-SEP-2003.

XX 28-FEB-2003; 2003WO-JP002335.

XX 28-FEB-2002; 2002JP-00053145.

XX (MEIJ ) MEIJ SEIKA KAISHA LTD.

PI Fujiaki K, Nagasawa H, Igarashi I, Suzuki H, Sugimoto C, Gen G;

PI Yu M, Tsuji N;

XX WPI; 2003-721752/68.

DR N-PSDB; ADB78971.

PT Tick chitinase and its encoded polynucleotide, applicable in developing

PT mite mediate infections e.g. piroplasmosis and Q fever.

XX Claim 1; Page 33-39; 47pp; Japanese.

XX The invention relates to a tick chitinase polypeptide, or its derivative  
 CC that has chitinase activity. Also disclosed is a vector containing the  
 CC polynucleotide, a transformant containing the polynucleotide, a process  
 CC for producing the polypeptide by culturing the transformant, drugs  
 CC containing the polypeptide, and a method for exterminating mites by  
 CC administering an effective dose of the polypeptide. The protein and its  
 CC encoded polynucleotide are applicable in developing diagnostics, agents  
 CC for exterminating mites, remedies or preventives, including vaccines, for  
 CC mite mediated infections e.g. piroplasmosis, Q fever or viral  
 CC encephalitis. The produced drugs are new and mostly likely less prone to

CC drug resistance and are safer. The current sequence represents the tick  
 CC chitinase amino acid sequence.

XX SQ Sequence 929 AA;

Query Match 70.3%; Score 45; DB 7; Length 929;  
 Best Local Similarity 61.5%; Pred. No. 30;  
 Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKONYLALVRELK 13  
 Db 175 DKONFVELVRELK 187

#### RESULT 14

ID ABB49221 standard; protein; 756 AA.

XX ABB49221;

DT 05-FEB-2002 (first entry)

DE Listeria monocytogenes protein #1925.

XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;

KM vitamin B12; bacterial infection; disease.

XX Listeria monocytogenes.

XX WO200177335-A2.

XX 18-OCT-2001.

XX 11-APR-2001; 2001WO-FR001118.

XX 11-APR-2000; 2000FR-00004629.

XX (INSP ) INST PASTEUR.

PI Buchrieser C, Frangeul L, Couve E, Rusznick C, Fsihi H, Dehoux P;

PI Dussanget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;

PI Daniels U, Goebel W, Kreft U, Kuhn M, Ng E, Vazquez-Boland JA;

PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;

PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;

PI Perez-Diaz J, Baguero F, Garcia Del Portillo F, Gomez-Lopez N;

PI Madueno E, De Pablo B, Wehland J, Kaerst U, Entian K, Hauf J;

XX Rose M, Voss H;

XX WPI; 2002-010914/01.

PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment

PT and prevention of Listeria and related bacterial infections, and related

PT polypeptides.

XX Claim 6; SEQ ID NO 1926; 192pp; French.  
 XX The present invention relates to the genome sequence of Listeria  
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
 CC it are useful for selecting probes and primers for detecting genes in L.  
 CC monocytogenes and related organisms, and for studying genetic  
 CC polymorphisms and other genomes. The present sequence is a protein  
 CC encoded by the genome sequence of the present invention. Proteins  
 CC expressed from the genome sequence are useful for raising specific  
 CC antibodies, identification of L. monocytogenes and related organisms, and  
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
 CC B12. The genome sequence and proteins encoded by it are also useful for  
 CC selecting compounds that regulate gene expression and cell replication  
 CC and modulate L. monocytogenes-related diseases. In addition, the genome  
 CC sequence and proteins encoded by it are useful in pharmaceutical and  
 CC vaccines compositions for the treatment or prevention of infections by L.  
 CC monocytogenes and related organisms. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX SQ Sequence 756 AA;  
 Query Match 64.1%; Score 41; DB 5; Length 756;  
 Best Local Similarity 46.2%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 DKONYLALVRELK 13  
 |||::|::|:  
 Db 198 DKQNFITLQDLR 210

SQ Sequence 1042 AA;  
 Query Match 64.1%; Score 41; DB 4; Length 1042;  
 Best Local Similarity 58.3%; Pred. No. 1.9e+02;  
 Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 KQNYLALVRELK 13  
 |||::|::|:  
 Db 645 KENYKALKEIK 656

Search completed: March 22, 2004, 06:51:41  
 Job time : 4.95275 secs

RESULT 15  
 ABG17202  
 ID ABG17202 standard; protein; 1042 AA.  
 XX AC ABG17202;  
 XX DT 18-FEB-2002 (first entry)  
 XX DE Novel human diagnostic protein #17193.  
 XX KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX KM food supplement; medical imaging; diagnostic; genetic disorder.  
 XX OS Homo sapiens.  
 XX PN WC200175067-A2.  
 XX PD 11-OCT-2001.  
 XX PF 30-MAR-2001; 2001MO-US008631.  
 XX PR 31-MAR-2000; 2000US-00540217.  
 XX PR 23-AUG-2000; 2000US-00649167.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Dmanac RT, Liu C, Tang YT;  
 XX DR N-PDSB; AAS81389.  
 XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 XX PT diagnostics, forensics, gene mapping, identification of mutations  
 XX PT responsible for genetic disorders or other traits and to assess  
 XX PT biodiversity.  
 XX PS Claim 20; SEQ ID NO 47561; 103pp; English.  
 XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 XX CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 XX CC and in recombinant production of (II). The polynucleotides are also used  
 XX CC in diagnostics as expressed sequence tags for identifying expressed  
 XX CC genes. (I) is useful in gene therapy techniques to restore normal  
 XX CC activity of (II) or to treat disease states involving (II). (II) is  
 XX CC useful for generating antibodies against it, detecting or quantitating a  
 XX CC polypeptide in tissue, as molecular weight markers and as a food  
 XX CC supplement. (II) and its binding partners are useful in medical imaging  
 XX CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 XX CC involving aberrant protein expression or biological activity. The  
 XX CC polypeptide and polynucleotide sequences have applications in  
 XX CC diagnostics, forensics, gene mapping, identification of mutations  
 XX CC responsible for genetic disorders or other traits to assess biodiversity  
 XX CC and to produce other types of data and products dependent on DNA and  
 XX CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
 XX CC amino acid sequences of the invention. Note: The sequence data for this  
 XX CC patent did not appear in the printed specification, but was obtained in  
 XX CC electronic format directly from WIPO at  
 XX CC ftp.wipo.int/pub/published\_pct\_sequences

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:59:34 ; Search time 2.65926 Seconds  
(without alignments)  
1265.926 Million cell updates/sec

Title: US-09-662-293-7  
Perfect score: 64  
Sequence: 1 DKONYALVRELK 13

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1049977 seqs, 258955339 residues

Total number of hits satisfying chosen parameters: 1049977

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubppaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubppaa/PCNUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubppaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubppaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	64	100.0	13	US-10-218-743-7	Sequence 7, Appli
2	64	100.0	536	US-10-218-743-21	Sequence 21, Appl
3	64	100.0	555	US-10-218-743-15	Sequence 15, Appl
4	64	100.0	555	US-10-218-743-18	Sequence 18, Appl
5	60	93.8	490	US-10-218-743-41	Sequence 41, Appl
6	60	93.8	509	US-10-218-743-35	Sequence 35, Appl
7	60	93.8	509	US-10-218-743-38	Sequence 38, Appl
8	46	71.9	617	US-10-369-493-6743	Sequence 6743, Ap
9	43	67.2	154	US-10-424-599-172346	Sequence 172346,
10	41	64.1	789	US-10-369-493-12804	Sequence 12804, A
11	40	62.5	84	US-10-369-493-23404	Sequence 23404, A
12	40	62.5	383	US-10-097-340-45	Sequence 45, Appl
13	40	62.5	383	US-10-095-027-270	Sequence 270, App
14	40	62.5	392	US-10-369-493-19361	Sequence 19361, A
15	40	62.5	635	US-10-369-493-16746	Sequence 16746, A

	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
US-10-218-743-7	39	38	37	37	37	37	37	37	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	35
Sequence 7, Appli	60.9	59.4	57.8	57.8	57.8	57.8	57.8	57.8	56.2	56.2	56.2	56.2	56.2	56.2	56.2	56.2	56.2	56.2	56.2	56.2	56.2	56.2	56.2	56.2	56.2	56.2	56.2	56.2	56.2	54.7
US-10-156-761-14515	407	987	331	407	407	407	407	407	184	184	185	12	293	293	293	293	293	293	358	358	358	358	358	358	358	358	358	358	358	244
Sequence 20271, A	US-10-369-493-20271	US-10-156-761-14515	US-10-128-714-8255	US-10-282-1224-63624	US-10-369-493-21317	US-10-369-493-5525	US-10-369-493-5525	US-10-369-493-5525	US-10-424-599-237039	US-10-369-493-20263	US-10-424-599-237039	US-10-062-920-40	US-10-062-920-40	US-10-062-920-40	US-10-062-920-40	US-10-062-920-40	US-10-062-920-40	US-10-062-920-40	US-10-062-920-40	US-10-062-920-40	US-10-062-920-40	US-10-062-920-40	US-10-062-920-40	US-10-062-920-40	US-10-062-920-40	US-10-062-920-40	US-10-062-920-40	US-10-062-920-40	US-10-062-920-40	US-10-062-920-40
Sequence 1515, A	Sequence 8255, Ap	Sequence 63624, A	Sequence 21317, A	Sequence 5036, Ap	Sequence 7036, Ap	Sequence 57720, A	Sequence 18943, A	Sequence 20260, A	Sequence 237039, A	Sequence 40, Appl	Sequence 40, Appl	Sequence 44, Appl	Sequence 44, Appl	Sequence 44, Appl	Sequence 44, Appl	Sequence 44, Appl	Sequence 44, Appl	Sequence 44, Appl	Sequence 44, Appl	Sequence 44, Appl	Sequence 44, Appl	Sequence 44, Appl	Sequence 44, Appl	Sequence 44, Appl	Sequence 44, Appl	Sequence 44, Appl	Sequence 44, Appl	Sequence 44, Appl	Sequence 44, Appl	Sequence 44, Appl
Sequence 153, App	Sequence 257, App	Sequence 30, Appl	Sequence 80, Appl	Sequence 75, Appl	Sequence 199, App	Sequence 182, App	Sequence 182, App	Sequence 182, App	Sequence 182, App	Sequence 182, App	Sequence 182, App	Sequence 182, App	Sequence 182, App	Sequence 182, App	Sequence 182, App	Sequence 182, App	Sequence 182, App	Sequence 182, App	Sequence 182, App	Sequence 182, App	Sequence 182, App	Sequence 182, App	Sequence 182, App	Sequence 182, App	Sequence 182, App	Sequence 182, App	Sequence 182, App	Sequence 182, App	Sequence 182, App	Sequence 182, App
Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App	Sequence 251698, App

RESULT 1  
US-10-218-743-7  
Sequence 7, Application US/10218743  
Publication No. US20030096779A1  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley Wu  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
TITLE OF INVENTION: AND USES THEREOF  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/10/218,743  
CURRENT FILING DATE: 2002-08-13  
PRIOR APPLICATION NUMBER: US/09/292,225  
PRIOR FILING DATE: 1999-04-15  
PRIOR APPLICATION NUMBER: 60/098,909  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/085,295  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/098,565  
PRIOR FILING DATE: 1998-04-17  
PRIOR APPLICATION NUMBER: 09/062,013  
PRIOR FILING DATE: 1998-04-17  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: Patent Ver. 2.0  
SEQ ID NO 7  
LENGTH: 13  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-10-218-743-7  
Query Match 100.0% Score 64; DB 14; Length 13;  
Best Local Similarity 100.0% Pred. NO. 6.9e-05;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DKONYALVRELK 13  
DB 1 DKONYALVRELK 13

```

RESULT 2
US-10-218-743-21
Sequence 21, Application US/10218743
Publication No. US20030096779A1
GENERAL INFORMATION:
APPLICANT: McCall, Catherine A.
APPLICANT: Hunter, Shirley Wu
APPLICANT: Weber, Eric R.
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: AL-2-C3
CURRENT APPLICATION NUMBER: US/10/218,743
CURRENT FILING DATE: 2002-08-13
PRIORITY APPLICATION NUMBER: US/09/292,225
PRIORITY FILING DATE: 1999-04-15
PRIORITY APPLICATION NUMBER: 60/098,909
PRIORITY FILING DATE: 1998-09-02
PRIORITY APPLICATION NUMBER: 60/085,295
PRIORITY FILING DATE: 1998-05-13
PRIORITY APPLICATION NUMBER: 60/098,565
PRIORITY FILING DATE: 1998-04-17
PRIORITY APPLICATION NUMBER: 60/098,565
PRIORITY FILING DATE: 1998-04-17
PRIORITY APPLICATION NUMBER: 09/062,013
PRIORITY FILING DATE: 1998-04-17
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 21
LENGTH: 536
TYPE: PRT
ORGANISM: Dermatophagoides farinae
US-10-218-743-21

Query Match 100.0%; Score 64; DB 14; Length 536;
Best Local Similarity 100.0%; Pred. No. 0.0042;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 1 DKONYLALVRELK 13
|||
Db 153 DKONYLALVRELK 165

RESULT 3
US-10-218-743-15
Sequence 15, Application US/10218743
Publication No. US20030096779A1
GENERAL INFORMATION:
APPLICANT: McCall, Catherine A.
APPLICANT: Hunter, Shirley Wu
APPLICANT: Weber, Eric R.
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: AL-2-C3
CURRENT APPLICATION NUMBER: US/10/218,743
CURRENT FILING DATE: 2002-08-13
PRIORITY APPLICATION NUMBER: US/09/292,225
PRIORITY FILING DATE: 1999-04-15
PRIORITY APPLICATION NUMBER: 60/098,509
PRIORITY FILING DATE: 1998-09-02
PRIORITY APPLICATION NUMBER: 60/085,295
PRIORITY FILING DATE: 1998-05-13
PRIORITY APPLICATION NUMBER: 60/098,565
PRIORITY FILING DATE: 1998-04-17
PRIORITY APPLICATION NUMBER: 09/062,013
PRIORITY FILING DATE: 1998-04-17
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 15
LENGTH: 555
TYPE: PRT
ORGANISM: Dermatophagoides farinae
US-10-218-743-15

```

```

QY      1 DKONYLATVRELK 13          100.0%; Score 64; DB 14; Length 555;
       |||||
Db      172 DKONYLATVRELK 184        Best Local Similarity 100.0%; Pred. No. 0.0043;
                                         Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0

RESULT 4
US-10-218-743-18
; Sequence 18, Application US/10218743
; Publication No. US20030096779A1
GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: AL-2-C3
CURRENT APPLICATION NUMBER: US/10/218,743
CURRENT FILING DATE: 2002-08-13
PRIOR APPLICATION NUMBER: US/09/292,225
PRIOR FILING DATE: 1999-04-15
PRIOR APPLICATION NUMBER: 60/098,909
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/085,295
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/098,565
PRIOR FILING DATE: 1998-04-17
PRIOR APPLICATION NUMBER: 09/062,013
PRIOR FILING DATE: 1998-04-17
NUMBER OF SEQ ID NOS: 49
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 18
LENGTH: 555
TYPE: PRT
ORGANISM: Dermatophagoides farinae
US-10-218-743-18

Query Match      100.0%; Score 64; DB 14; Length 555;
Best Local Similarity 100.0%; Pred. No. 0.0043;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY      1 DKONYLATVRELK 13
       |||||
Db      172 DKONYLATVRELK 184

RESULT 5
US-10-218-743-41
; Sequence 41, Application US/10218743
; Publication No. US20030096779A1
GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: AL-2-C3
CURRENT APPLICATION NUMBER: US/10/218,743
CURRENT FILING DATE: 2002-08-13
PRIOR APPLICATION NUMBER: US/09/292,225
PRIOR FILING DATE: 1999-04-15
PRIOR APPLICATION NUMBER: 60/098,909
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/085,295
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/098,565
PRIOR FILING DATE: 1998-04-17
PRIOR APPLICATION NUMBER: 09/062,013
PRIOR FILING DATE: 1998-04-17
NUMBER OF SEQ ID NOS: 49

```

SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 41  
LENGTH: 490  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-10-218-743-41

Query Match 93.8%; Score 60; DB 14; Length 490;  
Best Local Similarity 92.3%; Pred. No. 0.02;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKONYLALVRELK 13  
Db 153 DKONYLTLVRELK 165

RESULT 6  
US-10-218-743-35  
Sequence 35, Application US/10218743  
Publication No. US20030096779A1  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley Wu  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/10/218, 743  
CURRENT FILING DATE: 2002-08-13  
PRIOR APPLICATION NUMBER: US/09/292, 225  
PRIOR FILING DATE: 1999-04-15  
PRIOR APPLICATION NUMBER: 60/098, 909  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/085, 295  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/098, 565  
PRIOR FILING DATE: 1998-04-17  
PRIOR APPLICATION NUMBER: 09/062, 013  
PRIOR FILING DATE: 1998-04-17  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 35  
LENGTH: 509  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-10-218-743-35

Query Match 93.8%; Score 60; DB 14; Length 509;  
Best Local Similarity 92.3%; Pred. No. 0.021;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKONYLALVRELK 13  
Db 172 DKONYLTLVRELK 184

RESULT 7  
US-10-218-743-38  
Sequence 38, Application US/10218743  
Publication No. US20030096779A1  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley Wu  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/10/218, 743  
CURRENT FILING DATE: 2002-08-13  
PRIOR APPLICATION NUMBER: US/09/292, 225  
PRIOR FILING DATE: 1999-04-15  
PRIOR APPLICATION NUMBER: 60/098, 909  
PRIOR FILING DATE: 1998-09-02

PRIOR APPLICATION NUMBER: 60/085, 295  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/098, 565  
PRIOR FILING DATE: 1998-04-17  
PRIOR APPLICATION NUMBER: 09/062, 013  
PRIOR FILING DATE: 1998-04-17  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 38  
LENGTH: 509  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-10-218-743-38

Query Match 93.8%; Score 60; DB 14; Length 509;  
Best Local Similarity 92.3%; Pred. No. 0.021;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKONYLALVRELK 13  
Db 172 DKONYLTLVRELK 184

RESULT 8  
US-10-369-493-6743  
Sequence 6743, Application US/10369493  
Publication No. US2003023675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369, 493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360, 039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 6743  
LENGTH: 617  
TYPE: PRT  
ORGANISM: Caenorhabditis elegans  
US-10-369-493-6743

Query Match 71.9%; Score 46; DB 15; Length 617;  
Best Local Similarity 69.2%; Pred. No. 8.5;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 DKONYLALVRELK 13  
Db 186 DMANYVALVRELK 198

RESULT 9  
US-10-424-599-172346  
Sequence 172346, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424, 599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 172346  
LENGTH: 154

TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_126645C.1.pep  
US-10-424-599-172346

Query Match 67.2%; Score 43; DB 12; Length 154;  
Best Local Similarity 69.2%; Pred. No. 6.4;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 DKONYALVRELK 13  
DB 130 DKANHIALVRYLK 142

RESULT 10  
US-10-369-493-12804  
Sequence 12804, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 12804  
LENGTH: 789  
TYPE: PRT  
ORGANISM: Aspergillus nidulans  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(789)  
OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-12804

Query Match 64.1%; Score 41; DB 15; Length 789;  
Best Local Similarity 58.3%; Pred. No. 89;  
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 DKONYALVREL 12  
DB 217 DYENVSLVREM 228

RESULT 11  
US-10-369-493-23404  
Sequence 23404, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 23404  
LENGTH: 84  
TYPE: PRT

ORGANISM: Deinococcus radiodurans  
US-10-369-493-23404

Query Match 62.5%; Score 40; DB 15; Length 84;  
Best Local Similarity 66.7%; Pred. No. 11;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 KONYALVRELK 13  
DB 68 QOEYVALVEIK 79

RESULT 12  
US-10-097-340-45  
Sequence 45, Application US/10097340  
Publication No. US20030087250A1  
GENERAL INFORMATION:  
APPLICANT: JOHN MONAHAN  
APPLICANT: Manjula GANNANAVARAPU  
APPLICANT: Sebastian HOERSCH  
APPLICANT: Shubhangi KAMATKAR  
APPLICANT: Steve G. KOVATS  
APPLICANT: Rachel E. MEYERS  
APPLICANT: Michael MORRISSEY  
APPLICANT: Peter OLANDT  
APPLICANT: Ami SEN  
APPLICANT: Peter VERIBY  
APPLICANT: Gordon B. MILLS  
APPLICANT: Robert C. BAST, Jr.  
APPLICANT: Karen LU  
APPLICANT: Rosemarie SCHMANDT  
APPLICANT: Xumei ZHAO  
APPLICANT: Karen GLATT  
TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,  
TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer  
FILE REFERENCE: MRI-030  
CURRENT APPLICATION NUMBER: US/10/097,340  
CURRENT FILING DATE: 2002-03-14  
PRIOR APPLICATION NUMBER: 60/276,025  
PRIOR FILING DATE: 2001-03-14  
PRIOR APPLICATION NUMBER: 60/325,149  
PRIOR FILING DATE: 2001-09-26  
PRIOR APPLICATION NUMBER: 60/325,149  
PRIOR FILING DATE: 2001-03-14  
PRIOR APPLICATION NUMBER: 60/276,026  
PRIOR FILING DATE: 2001-09-26  
PRIOR APPLICATION NUMBER: 60/324,967  
PRIOR FILING DATE: 2001/09/26  
PRIOR APPLICATION NUMBER: 60/311,732  
PRIOR FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: 60/325,102  
PRIOR FILING DATE: 2001-09-26  
PRIOR APPLICATION NUMBER: 60/323,580  
PRIOR FILING DATE: 2001-09-19  
NUMBER OF SEQ ID NOS: 363  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 45  
LENGTH: 383  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-097-340-45

Query Match 62.5%; Score 40; DB 14; Length 383;  
Best Local Similarity 46.2%; Pred. No. 60;  
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 DKONYALVRELK 13  
DB 146 DKQHFTLIREMK 158

RESULT 13  
US-10-295-027-270  
Sequence 270, Application US/10295027  
Publication No. US20030232350A1

```

; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezl, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 270
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-295-027-270

Query Match          62.5%; Score 40; DB 15; Length 383;
Best Local Similarity 46.2%; Pred. No. 60;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY      1 DKONYLALVRELK 13
Db      146 DKQHTLLIKMK 158

RESULT 14
US-10-369-493-19361
; Sequence 19361, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 19361

```

```

; LENGTH: 392
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
; US-10-369-493-19361

Query Match          62.5%; Score 40; DB 15; Length 392;
Best Local Similarity 53.8%; Pred. No. 62;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY      1 DKONYLALVRELK 13
Db      160 DKQNTLLMOEPR 172

RESULT 15
US-10-369-493-16746
; Sequence 16746, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 16746
; LENGTH: 635
; TYPE: PRT
; ORGANISM: Bacillus thuringiensis
; US-10-369-493-16746

Query Match          62.5%; Score 40; DB 15; Length 635;
Best Local Similarity 53.8%; Pred. No. 11e+02;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY      1 DKONYLALVRELK 13
Db      225 DKQNTLLMOEPR 237

Search completed: March 22, 2004, 07:45:49
Job time : 3.65926 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:42:54 ; Search time 1.0478 Seconds  
(without alignments)  
640.518 Million cell updates/sec

Title: US-09-662-293-7  
Perfect score: 64  
Sequence: 1 DKONYALVREIK 13

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	64	100.0	13	US-09-292-225-7	Sequence 7, Appl
2	64	100.0	536	US-09-292-225-21	Sequence 21, Appl
3	64	100.0	555	US-09-292-225-15	Sequence 15, Appl
4	64	100.0	555	US-09-292-225-18	Sequence 18, Appl
5	60	93.8	490	US-09-292-225-41	Sequence 41, Appl
6	60	93.8	509	US-09-292-225-35	Sequence 35, Appl
7	60	93.8	509	US-09-292-225-38	Sequence 38, Appl
8	40	62.5	511	US-09-489-039A-12116	Sequence 12116, A
9	39	60.9	289	US-08-481-985B-79	Sequence 79, Appl
10	39	60.9	289	US-08-481-985B-79	Sequence 79, Appl
11	39	60.9	289	US-08-370-476-79	Sequence 79, Appl
12	38	59.4	442	US-09-052-778-2	Sequence 2, Appl
13	37	57.8	333	US-09-107-532A-4886	Sequence 4886, Ap
14	37	57.8	675	US-09-134-001C-4356	Sequence 4356, A
15	36	56.2	191	US-09-258-991A-30217	Sequence 30217, A
16	36	56.2	293	US-09-660-587-40	Sequence 40, Appl
17	36	56.2	293	US-09-314-701-44	Sequence 44, Appl
18	36	56.2	293	US-09-811-007A-40	Sequence 40, Appl
19	36	56.2	554	US-08-524-051-2	Sequence 2, Appl
20	36	56.2	554	US-09-052-778-16	Sequence 16, Appl
21	36	56.2	752	US-08-420-235B-21	Sequence 21, Appl
22	36	56.2	752	US-08-793-624-21	Sequence 21, Appl
23	36	56.2	752	PCT-US95-10194-21	Sequence 21, Appl
24	35	54.7	16	US-09-171-705-21	Sequence 21, Appl
25	35	54.7	144	US-08-133-979A-21	Sequence 21, Appl
26	35	54.7	144	US-08-436-890-21	Sequence 21, Appl
27	35	54.7	144	US-08-451-213-21	Sequence 21, Appl

28	35	54.7	290	2	US-08-484-905-80	Sequence 80, Appl
29	35	54.7	290	3	US-08-481-985B-80	Sequence 80, Appl
30	35	54.7	280	3	US-08-370-476-80	Sequence 80, Appl
31	35	54.7	293	4	US-09-314-701-54	Sequence 54, Appl
32	35	54.7	297	4	US-09-314-701-14	Sequence 14, Appl
33	35	54.7	301	2	US-08-484-905-77	Sequence 77, Appl
34	35	54.7	301	2	US-08-481-985B-77	Sequence 77, Appl
35	35	54.7	301	3	US-08-370-476-77	Sequence 77, Appl
36	35	54.7	308	2	US-08-484-905-72	Sequence 72, Appl
37	35	54.7	308	3	US-08-481-985B-72	Sequence 72, Appl
38	35	54.7	308	3	US-08-370-476-72	Sequence 72, Appl
39	35	54.7	310	2	US-08-484-905-66	Sequence 66, Appl
40	35	54.7	310	3	US-08-481-985B-66	Sequence 66, Appl
41	35	54.7	310	3	US-08-370-476-66	Sequence 66, Appl
42	35	54.7	326	4	US-09-328-352-5506	Sequence 5506, Ap
43	35	54.7	331	4	US-09-134-001C-5254	Sequence 5254, Ap
44	35	54.7	344	4	US-09-134-001C-3524	Sequence 3524, Ap
45	35	54.7	445	4	US-09-543-681A-4337	Sequence 4337, Ap

## ALIGNMENTS

```
RESULT 1
US-09-292-225-7
; Sequence 7, Application US/09292225
; Patent No. 6455686
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; EARLIER FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: 60/098,909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085,295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098,565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062,013
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 7
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-09-292-225-7
Query Match 100.0%; Score 64; DB 4; Length 13;
Best Local Similarity 100.0%; Pred. No. 4e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cy 1 DKONYALVREIK 13
Db 1 DKONYALVREIK 13
RESULT 2
US-09-292-225-21
; Sequence 21, Application US/09292225
; Patent No. 6455686
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
```



```

; CURRENT FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: 60/098,909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085,295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098,565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062,013
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO: 21
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-09-292-225-21

```

```

Query Match      100.0%; Score 64; DB 4; Length 536;
Best Local Similarity 100.0%; Pred. No. 0.0021;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```
Qy      1 DKONYALVRELK 13
Db      153 DKONYALVRELK 165

```

```

RESULT 3
US-09-292-225-15
; Sequence 15, Application US/09292225
; Patent No. 6455686
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; CURRENT FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: 60/098,909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085,295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098,565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062,013
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO: 15
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-09-292-225-15

```

```

; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; CURRENT FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: 60/098,909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085,295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098,565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062,013
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO: 15
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-09-292-225-15

```

```

Query Match      100.0%; Score 64; DB 4; Length 555;
Best Local Similarity 100.0%; Pred. No. 0.0021;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```
Qy      1 DKONYALVRELK 13
Db      172 DKONYALVRELK 184

```

```

RESULT 4
US-09-292-225-18
; Sequence 18, Application US/09292225
; Patent No. 6455686
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
US-09-292-225-18

```

```

; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; CURRENT FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: 60/098,909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085,295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098,565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062,013
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO: 18
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-09-292-225-18

```

```

Query Match      100.0%; Score 64; DB 4; Length 555;
Best Local Similarity 100.0%; Pred. No. 0.0021;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```
Qy      1 DKONYALVRELK 13
Db      172 DKONYALVRELK 184

```

```

RESULT 5
US-09-292-225-41
; Sequence 41, Application US/09292225
; Patent No. 6455686
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; CURRENT FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: 60/098,909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085,295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098,565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062,013
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO: 41
; LENGTH: 490
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-09-292-225-41

```

```

Query Match      93.8%; Score 60; DB 4; Length 490;
Best Local Similarity 92.3%; Pred. No. 0.0095;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```
Qy      1 DKONYALVRELK 13
Db      153 DKONYALVRELK 165

```

```

RESULT 6
US-09-292-225-35
; Sequence 35, Application US/09292225
; Patent No. 6455686
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
US-09-292-225-35

```

APPLICANT: Hunter, Shirley Wu  
APPLICANT: Weber, Eric R.  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/09/292,225  
CURRENT FILING DATE: 1999-04-15  
EARLIER APPLICATION NUMBER: 60/098,909  
EARLIER FILING DATE: 1998-09-02  
EARLIER APPLICATION NUMBER: 60/085,295  
EARLIER FILING DATE: 1998-05-13  
EARLIER APPLICATION NUMBER: 60/098,565  
EARLIER FILING DATE: 1998-04-17  
EARLIER APPLICATION NUMBER: 09/062,013  
EARLIER FILING DATE: 1998-04-17  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 35  
LENGTH: 509  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-09-292-225-35

Query Match 93.8%; Score 60; DB 4; Length 509;  
Best Local Similarity 92.3%; Pred. No. 0.0099;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 DKONYLALVRELK 13  
Db 172 DKONYLALVRELK 184

RESULT 7  
US-09-292-225-38  
Sequence 38, Application US/09292225  
Patent No. 6455686  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley Wu  
APPLICANT: Weber, Eric R.  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/09/292,225  
CURRENT FILING DATE: 1999-04-15  
EARLIER APPLICATION NUMBER: 60/098,909  
EARLIER FILING DATE: 1998-09-02  
EARLIER APPLICATION NUMBER: 60/085,295  
EARLIER FILING DATE: 1998-05-13  
EARLIER APPLICATION NUMBER: 60/098,565  
EARLIER FILING DATE: 1998-04-17  
EARLIER APPLICATION NUMBER: 09/062,013  
EARLIER FILING DATE: 1998-04-17  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 38  
LENGTH: 509  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-09-292-225-38

Query Match 93.8%; Score 60; DB 4; Length 509;  
Best Local Similarity 92.3%; Pred. No. 0.0099;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 DKONYLALVRELK 13  
Db 172 DKONYLALVRELK 184

RESULT 8  
US-09-489-039A-12116  
Sequence 12116, Application US/09489039A

Patent No. 6610836  
GENERAL INFORMATION:  
APPLICANT: Gary Breton et. al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
FILE REFERENCE: 2709,2004001  
CURRENT APPLICATION NUMBER: US/09/489,039A  
CURRENT FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: US 60/117,747  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 14342  
SEQ ID NO 12116  
LENGTH: 611  
TYPE: PRT  
ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-12116

Query Match 62.5%; Score 40; DB 4; Length 611;  
Best Local Similarity 50.0%; Pred. No. 40;  
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 1 DKONYLALVREL 12  
Db 448 DKEDYALIEQM 459

RESULT 9  
US-08-484-905-79  
Sequence 79, Application US/08484905  
Patent No. 5976551  
GENERAL INFORMATION:  
APPLICANT: Mottez, Estelle  
APPLICANT: Abastado, Jean-Pierre  
APPLICANT: Kourilsky, Philippe  
TITLE OF INVENTION: An Altered Major Histocompatibility  
TITLE OF INVENTION: Complex(MHC) Determinant and Methods for Using the  
NUMBER OF SEQUENCES: 127  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESSEE: Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS-/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,905  
FILING DATE: 07-JUNE-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/801,818  
FILING DATE: 05-DEC-1991  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/792,473  
FILING DATE: 15-NOV-1991  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Pocter, Jane E. R.  
REGISTRATION NUMBER: 33,332  
REFERENCE/DOCKET NUMBER: 03495,0106-03000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 79:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 289 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-484-905-79

Query Match  
Best Local Similarity 53.8%; Score 39; DB 2; Length 289;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 DKONYIALVRELK 13  
|:|:|:|:|:  
Db 119 DGQDYIALMEDLK 131

RESULT 10  
US-08-481-985B-79  
Sequence 79, Application US/08481985B  
Patent No. 6011146

GENERAL INFORMATION:  
APPLICANT: Mottez, Estelle  
APPLICANT: Abastado, Jean-Pierre  
APPLICANT: Kourilsky, Philippe  
TITLE OF INVENTION: Altered Major Histocompatibility Complex  
TITLE OF INVENTION:  
NUMBER OF SEQUENCES: 148  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESSEE: Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
ZIP: 20005-3315

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/481,985B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/801,818  
FILING DATE: 05-DEC-1991  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/792,473  
FILING DATE: 15-NOV-1991

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03495.0106-04000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 79:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 289 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-481-985B-79

Query Match  
Best Local Similarity 60.9%; Score 39; DB 3; Length 289;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 DKONYIALVRELK 13  
|:|:|:|:|:  
Db 119 DGQDYIALMEDLK 131

RESULT 11

US-08-370-476-79  
Sequence 79, Application US/08370476  
Patent No. 6153408

GENERAL INFORMATION:  
APPLICANT: Mottez, Estelle  
APPLICANT: Abastado, Jean-Pierre  
APPLICANT: Kourilsky, Philippe  
APPLICANT: Lome, Yu-Chun  
APPLICANT: Ojcius, David  
APPLICANT: Casrouge, Armand  
TITLE OF INVENTION: Altered Major Histocompatibility Complex  
TITLE OF INVENTION:  
NUMBER OF SEQUENCES: 127  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESSEE: Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
ZIP: 20005-3315

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/370,476  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/117,575  
FILING DATE: 07-SEP-1993  
APPLICATION NUMBER: US 08/072,787  
FILING DATE: 06-JUN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/801,818  
FILING DATE: 05-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/792,473  
FILING DATE: 15-NOV-1991

ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 05243.0001-01000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 79:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 289 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-370-476-79

Query Match  
Best Local Similarity 60.9%; Score 39; DB 3; Length 289;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 DKONYIALVRELK 13  
|:|:|:|:|:  
Db 119 DGQDYIALMEDLK 131

RESULT 12  
US-09-052-778-2  
Sequence 2, Application US/09052778A  
Patent No. 6060590

GENERAL INFORMATION:  
APPLICANT: Bryant, Peter J.  
APPLICANT: Kawamura, Kazuo  
TITLE OF INVENTION: CHITINASE RELATED PROTEINS AND METHODS  
TITLE OF INVENTION: OF USE

FILE REFERENCE: 07306/015001  
 CURRENT APPLICATION NUMBER: US/09/052,778A  
 CURRENT FILING DATE: 1998-03-31  
 NUMBER OF SEQ ID NOS: 16  
 SOFTWARE: FASTSEQ for Windows Version 4.0  
 SEQ ID NO 2  
 LENGTH: 442  
 TYPE: PRT  
 ORGANISM: Drosophila melanogaster  
 US-09-052-778-2

Query Match 59.4%; Score 38; DB 3; Length 442;  
 Best Local Similarity 66.7%; Pred. No. 64;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 KONYALVRELK 13  
 : : : : :  
 : : : : :  
 Db 194 KEAFTALVRELK 205

RESULT 13  
 US-09-107-532A-4886

; Sequence 4886, Application US/09107532A  
 ; Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS: GENOME THERAPEUTICS CORPORATION

ADDRESSEE: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02154

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

FILING DATE: 30-Jun-1998

APPLICATION NUMBER: US/09/107,532A

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 4886:

SEQUENCE CHARACTERISTICS:

LENGTH: 333 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...333

SEQUENCE DESCRIPTION: SEQ ID NO: 4886:

US-09-107-532A-4886

Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 KONYALVREL 12  
 : : : : :  
 : : : : :  
 Db 155 KQOYLPLVREL 165

RESULT 14  
 US-09-134-001C-4356

; Sequence 4356, Application US/09134001C  
 ; Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 4356

LENGTH: 675

TYPE: PRT

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4356

Query Match 57.8%; Score 37; DB 4; Length 675;  
 Best Local Similarity 66.7%; Pred. No. 1,5e+02;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 KONYALVRELK 13  
 : : : : :  
 : : : : :  
 Db 135 KINYQALVRELK 146

RESULT 15  
 US-09-252-991A-30217

; Sequence 30217, Application US/09252991A  
 ; Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196,136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 30217

LENGTH: 191

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-30217

Query Match 56.2%; Score 36; DB 4; Length 191;  
 Best Local Similarity 66.7%; Pred. No. 60;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 DKONYALVREL 12  
 : : : : :  
 : : : : :  
 Db 32 DEQALVREL 43

Search completed: March 22, 2004, 07:03:56  
 Job time: 2.0478 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:40:28 ; Search time 1.10561 seconds  
(without alignments)

1479.047 Million cell updates/sec

Title: US-09-662-293-10  
Perfect score: 83  
Sequence: 1 DKLVGVFPYGRAXSIE 17

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61	73.5	499	2	S52422 chitinase (EC 3.2.
2	61	73.5	499	2	S04856 chitinase (EC 3.2.
3	58	69.9	554	2	A56596 chitinase (EC 3.2.
4	56	67.5	633	2	T24898 hypothetical prote
5	56	67.5	699	2	A38368 chitinase (EC 3.2.
6	55	66.3	572	2	G84238 hypothetical prote
7	55	66.3	599	2	D83764 chitinase BH0916 [
8	53	63.9	232	2	H69868 chitinase homolog
9	52	62.7	378	2	S51591 chitinase (EC 3.2.
10	52	62.7	1635	2	T147075 chitinase (EC 3.2.
11	51	61.4	457	2	JC6551 chitinase (EC 3.2.
12	51	61.4	654	2	I38605 oviductal glycopro
13	51	61.4	1484	2	T29275 hypothetical prote
14	50	60.2	383	2	S51327 heparin-binding gl
15	50	60.2	427	2	UC4565 chitinase (EC 3.2.
16	49	59.0	383	2	A49562 cartilage glycopro
17	49	59.0	385	2	A44102 di-N-acetylchitobi
18	49	59.0	413	2	JC2135 chitinase (EC 3.2.
19	49	59.0	424	2	S47133 chitinase (EC 3.2.
20	49	59.0	483	2	A53918 chitinase (EC 3.2.
21	49	59.0	525	2	T44445 chitinase (EC 3.2.
22	48	57.8	609	2	T42073 di-N-acetylchitobi
23	48	57.8	610	2	TH0573 chitinase (EC 3.2.
24	48	57.8	2025	2	TC03864 hypothetical prote
25	48	57.8	423	2	UC1975 chitinase (EC 3.2.
26	47	56.6	423	2	UC1975 chitinase (EC 3.2.
27	47	56.6	423	2	S51369 chitinase I precur
28	47	56.6	424	2	S68121 chitinase I precur
29	47	56.6	537	2	S57197 oviduct-specific g

30	47	56.6	539	2	I46470 estrogen dependent
31	46	55.4	558	2	T30418 chitinase (EC 3.2.
32	46	55.4	765	2	T35719 chitinase - Strept
33	45	54.2	799	2	PC4106 chitinase (EC 3.2.
34	45	54.2	820	2	A40633 chitinase (EC 3.2.
35	44	53.0	546	2	F84238 chitinase [impor
36	44	53.0	617	2	T15408 hypothetical prote
37	43	51.8	185	2	H86887 transcription anti
38	43	51.8	405	2	S61551 breast-regressing
39	43	51.8	465	2	E83449 conserved hypotet
40	42	50.6	216	2	S21337 genome polypept
41	42	50.6	248	2	T04758 hypothetical prote
42	42	50.6	344	2	H70030 conserved hypotet
43	42	50.6	426	2	D83936 hypothetical prote
44	42	50.6	429	2	B97238 protein containi
45	42	50.6	719	2	B86490 F28L22.6 protein -

## ALIGNMENTS

RESULT 1  
S52422  
chitinase (EC 3.2.1.14) B precursor - Serratia marcescens (strain BJL200)  
C/Species: Serratia marcescens  
A/Variety: strain BJL200  
C/Date: 01-Aug-1995 #sequence\_revision 03-Nov-1995 #ext\_change 08-Oct-1999  
C/Accession: S52422  
R/Bruberg, M.B.; Eklund, V.G.H.; Haandrikman, A.J.; Venema, G.; Nes, I.F.  
Microbiology 141, 123-131, 1995  
A/Title: Chitinase B from Serratia marcescens BJL200 is exported to the periplasm without  
A/Reference number: S52422; PMID:95202070; PMID:7894703  
A/Accession: S52422  
A/Molecule type: DNA  
A/Residues: 1-499 <BRU>  
A/Cross-references: EMBL:Z36295; NID:g677860; PIDN:CAA85292.1; PID:g677861  
A/Experimental source: strain BJL200  
C/Genetics:  
A/Gene: chb  
C/Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 73.5%; Score 61; DB 2; Length 499;  
Best Local Similarity 91.7%; Pred. No. 0.0074;  
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KLVGVFPYGRA 13  
DB 284 KLVGVFPYGRA 295

RESULT 2  
S04856  
chitinase (EC 3.2.1.14) B precursor - Serratia marcescens (strain QMB1466)  
C/Species: Serratia marcescens  
A/Variety: strain QMB1466  
C/Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #ext\_change 08-Oct-1999  
C/Accession: S04856  
R/Harpster, M.H.; Dunsin, P.  
Nucleic Acids Res. 17, 5395, 1989  
A/Title: Nucleotide sequence of the chitinase B gene of Serratia marcescens QMB1466.  
A/Reference number: S04856; PMID:89345110; PMID:268886  
A/Accession: S04856  
A/Molecule type: DNA  
A/Residues: 1-499 <HAR>  
A/Cross-references: EMBL:X15208; NID:g47227; PIDN:CAA33278.1; PID:g47228  
A/Experimental source: strain QMB1466  
C/Genetics:  
A/Gene: chb  
C/Keywords: glycosidase; hydrolase; polysaccharide degradation  
F/42-499/Product: chitinase B #status predicted <MAT>  
Query Match 73.5%; Score 61; DB 2; Length 499;  
Best Local Similarity 91.7%; Pred. No. 0.0074;

Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KLVWGVPPYGRA 13  
||:|||||||  
Db 284 KLVWGVPPYGRA 295

## RESULT 3

A56596  
chitinase (EC 3.2.1.14) - tobacco hornworm

C:Species: Manduca sexta (tobacco hornworm)

C>Date: 11-Aug-1995 #sequence\_revision 11-Aug-1995 #text\_change 29-Jan-1999

C:Accession: A56596

R:Kramer, K.J.; Corpuz, L.; Choi, H.K.; Muthukrishnan, S.

Insect Biochem. Mol. Biol. 23, 691-701, 1993

A:Title: Sequence of a cDNA and expression of the gene encoding epidermal and gut chitin

A:Reference number: A56596; PMID:8353525

A:Accession: A56596

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-354 <NR>

A:Cross-references: GB:U02270; GB:S64757; NID:g406048; PID:g406049

A:Experimental source: larvae

A:Note: sequence extracted from NCBI backbone (NCBIN:136417, NCBIPI:136418)

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 69.9%; Score 58; DB 2; Length 554;  
Best Local Similarity 56.2%; Pred. No. 0.028;  
Matches 9; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKLVMGVPPYGRAXSI 16  
:||||:||||: :  
Db 261 NKLWGVIPYGRSFTL 276

RESULT 4  
T24898  
hypothetical protein T13H5.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T24898

R:Lightning, J.

Submitted to the EMBL Data Library, October 1995

A:Reference number: Z19950

A:Accession: T24898

A:Status: preliminary; translated from GB/EMBL/DDJ

A:Molecule type: DNA

A:Residues: 1-633 <WIL>

A:Cross-references: EMBL:Z66524; PINN:CAA91419.1; GSPDB:GN00020; CESP:T13H5.3

A:Experimental source: clone T13H5

A:Gene: CESP:T13H5.3

A:Map position: 2

A:introns: 27/1; 57/3; 93/3; 126/2; 251/2; 274/3; 457/3; 578/1

Query Match 67.5%; Score 56; DB 2; Length 633;  
Best Local Similarity 83.3%; Pred. No. 0.073;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKLVMGVPPYGR 12  
||:|||||||  
Db 273 DKINMGVPPYGR 284

RESULT 5  
A38368  
chitinase (EC 3.2.1.14) precursor - Bacillus circulans

C:Species: Bacillus circulans

C>Date: 28-Jun-1991 #sequence\_revision 28-Jun-1991 #text\_change 15-Oct-1999

C:Accession: A38368

R:Matanabe, T.; Suzuki, K.; Oyanagi, W.; Ohnishi, K.; Tanaka, H.

J. Biol. Chem. 265, 15659-15665, 1990

A:Title: Gene cloning of chitinase A1 from Bacillus circulans WL-12 revealed its evolution

A:Reference number: A38368; PMID:9036876; PMID:2203782

A:Accession: A38368

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-699 <MAT>

A:Cross-references: GB:M57601; GB:J05599; NID:g1066341; PINN:AAA81528.1; PID:g142688

C:Superfamily: fibronectin type III repeat homology

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 67.5%; Score 56; DB 2; Length 699;  
Best Local Similarity 90.9%; Pred. No. 0.081;  
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KLVWGVPPYGR 12  
||:|||||||  
Db 330 KLVWGVPPYGR 340

RESULT 6  
G84238  
hypothetical protein Vng0818c [imported] - Halobacterium sp. NRC-1

C:Species: Halobacterium sp. NRC-1

C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C:Accession: G84238

R:Ng, W.V.; Kennedy, S.P.; Mahatras, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Laeky, S.

; Leitnauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Kough, D.W.; Maddocks, D.G.; Jablor

Jung, K.H.; Alam, M.; Freltas, T.

Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Edhardt, H.; Lowe, T.M.; Li

A:Title: Genome sequence of Halobacterium species NRC-1.

A:Reference number: A84160; PMID:20504483; PMID:11016950

A:Accession: G84238

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-572 <STO>

A:Cross-references: GB:AE004437; NID:g10580387; PINN:AA019275.1; GSPDB:GN00138

C:Genetics:

A:Gene: VNG0818C

Query Match 66.3%; Score 55; DB 2; Length 572;  
Best Local Similarity 60.0%; Pred. No. 0.099;  
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 LVWGVPPYGRAXSI 17  
||:|||||||  
Db 433 LVWGVPPYGRGNV 447

RESULT 7  
D83764  
chitinase BH0916 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001

C:Accession: D83764

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hiran

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; PMID:20512582; PMID:11058132

A:Accession: D83764

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-599 <STO>

A:Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PINN:BA04635.1; GSPDB:GN001

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH0916

Query Match 66.3%; Score 55; DB 2; Length 599;  
Best Local Similarity 71.4%; Pred. No. 0.1;  
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 KLVWGVPPYGRAXS 15  
||:|||||||

Db 325 KLVGMGPFYGRMS 338

# RESULT 8

H69868

chitinase homolog ykvo - *Bacillus subtilis*

C/Species: *Bacillus subtilis*  
C/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999

C/Accession: H69868

R/Kunet, F.; Ogawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berton  
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997

A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Galleg  
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois  
A.; Authors: Lamber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue  
Y., M.; Ogawa, K.; Ogawara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon  
A.; Authors: Schlicht, S.; Schreier, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron  
akuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpeira, P.; Tognoni, A.; Tostato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K.  
A/Authors: Yoshikawa, H.F.; Zimstein, E.; Yoshikawa, H.; Danchin, A.

A/Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A/Reference number: A69580; WUID:98044033; PMID:9384377

A/Accession: H69868

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-232 <KUN>

A/Cross-references: GB:299111; GB:AL009126; NID:92633699; PIDN:CAM13252.1; PID:el184869;

A/Experimental source: strain 168

C/Genetics:

A/Genes: ykvo

Query Match 63.9%; Score 53; DB 2; Length 232;

Best Local Similarity 81.8%; Pred. No. 0.086;

Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 KLVGMGPFYGR 12

Db 200 KLVGMGPFYGR 210

# RESULT 9

S51591

chitinase (EC 3.2.1.14) / lysozyme (EC 3.2.1.17) PZ precursor, pathogenesis-related - cc

C/Species: *Nicotiana tabacum* (common tobacco)

C/Date: 15-Jul-1995 #sequence\_revision 08-Sep-1995 #text\_change 22-Jun-1999

C/Accession: S51591; S51632; S43119

R/Reitz, T.; Second, S.; Kaufmann, S.; Geofroy, P.; Prasad, V.; Brunner, F.; Fritig, E

Mol. Gen. Genet. 245, 246-254, 1994

A/Title: Molecular characterization of a novel tobacco pathogenesis-related (PR) protein

A/Reference number: S51591; WUID:95115673; PMID:7816033

A/Accession: S51591

A/Molecule type: mRNA

A/Residues: 1-378 <HEI>

A/Cross-references: EMBL:X78325; NID:9467688; PIDN:CA55128.1; PID:9467689

A/Experimental source: cv. Samsun NN

A/Accession: S51632

A/Molecule type: protein

A/Residues: 31-36;87-112;252-275;282-305;337-371 <HEM>

C/Suprafamily: Streptomyces chitinase chi40

C/Keywords: glycosidase; hydrolase; polysaccharide degradation

F/1-25/Domain: signal sequence #status predicted <SIG>

F/26-371/Product: chitinase/lysozyme PZ #status experimental <MAT>

F/371-378/Domain: carboxyl-terminal propeptide #status predicted <PRO>

Query Match 62.7%; Score 52; DB 2; Length 378;

Best Local Similarity 75.0%; Pred. No. 0.22;

Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 KLVGMGPFYGR 13

Db 2 KLVGMGPFYGR 13

Db 252 KLVGMGPFYGR 263

# RESULT 10

T14075

chitinase (EC 3.2.1.14) - yellow fever mosquito

C/Species: *Aedes aegypti* (yellow fever mosquito)

C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C/Accession: T14075

R/de la Vega, H.; Specht, C.A.; Liu, Y.; Robbins, P.W.

Insect Mol. Biol. 7, 233-239, 1997

A/Title: Chitinases are a multi-gene family in *Aedes*, *Anopheles*, and *Drosophila*.

A/Reference number: 217872

A/Accession: T14075

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1635 <DEL>

A/Cross-references: EMBL:AF026492; NID:92564720; PID:92564721; PIDN:AA81850.1

C/Genetics:

A/Genes: CHIT2

A/Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 62.7%; Score 52; DB 2; Length 1635;

Best Local Similarity 60.0%; Pred. No. 1;

Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 KLVGMGPFYGR 16

Db 905 KLVGMGPFYGR 919

# RESULT 11

JC6551

chitinase (EC 3.2.1.14) precursor - *Leishmania donovani*

C/Species: *Leishmania donovani*

C/Date: 05-Dec-1998 #sequence\_revision 05-Dec-1998 #text\_change 07-May-1999

C/Accession: JC6551

R/Shakarian, A.M.; Dwyer, D.M.

Gene 208, 315-322, 1998

A/Title: The *ld* Chit1 gene encodes the secretory chitinase of the human pathogen *Leishman*

A/Reference number: JC6551; WUID:98201628; PMID:9524285

A/Accession: JC6551

A/Molecule type: DNA

A/Residues: 1-457 <SHA>

A/Cross-references: DDBJ:AF009354

A/Comment: This enzyme cleaves the beta-1-4 linkage between N-acetyl-glucosamine residues

C/Genetics:

A/Genes: chit1

C/Keywords: glycoprotein; glycosidase; glycosyltransferase; hexosyltransferase; hydrolase

F/1-19/Domain: signal sequence #status predicted <SIG>

F/20-457/Product: chitinase #status predicted <MAT>

F/384/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 61.4%; Score 51; DB 2; Length 457;

Best Local Similarity 72.7%; Pred. No. 0.4;

Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 KLVGMGPFYGR 12

Db 309 KLVGMGPFYGR 319

# RESULT 12

I38605

oviductal glycoprotein - human

C/Species: *Homo sapiens* (man)

C/Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 29-May-1998

C/Accession: I38605

R/Arias, E.B.; Verhage, H.G.; Jaffe, R.C.

Biol. Reprod. 51, 685-694, 1994

A/Title: Complementary deoxyribonucleic acid cloning and molecular characterization of an

A/Reference number: I38605; WUID:95119256; PMID:7819450

A:Accession: J38605  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-654 <RES>  
A:Cross-references: EMBL:U09550; NID:G529147; PID:G529148

Query Match 61.4%; Score 51; DB 2; Length 654;  
Best Local Similarity 66.7%; Pred. No. 0.59;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKLVGVPFYGR 12  
:|||||:  
Db 254 EKLWGIPTFYGR 265

## RESULT 13

T9275

hypothetical protein T01C4.1 - *Caenorhabditis elegans*C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T9275

R:Graves, T.; Woldmann, P.

submitted to the EMBL Data Library, September 1996

A:Description: The sequence of C. elegans cosmid T01C4.

A:Reference number: Z20599

A:Accession: T9275

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1484 &lt;GRA&gt;

A:Cross-references: EMBL:U70858; PIDN:AAB09177.1; GSPDB:GN0023; CESP:T01C4.1

A:Experimental source: strain Bristol N2; clone T01C4

C:Genetics:

A:Gene: CESP:T01C4.1

A:Map position: 5

A:introns: 60/1; 179/1; 219/3; 392/3; 481/3; 549/2; 594/2; 736/3; 843/3; 1056/2; 1444/3

Query Match 61.4%; Score 51; DB 2; Length 1484;  
Best Local Similarity 81.8%; Pred. No. 1.4;

Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 KLVGVPFYGR 12  
:|||||:  
Db 1199 QLVGVPFYGR 1209

## RESULT 14

S51327

heparin-binding glycoprotein 38K - pig

C:Species: *Sus scrofa domestica* (domestic pig)

C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 22-Jun-1999

C:Accession: S51327

R:Shackleton, L.M.; Mann, D.M.; Mills, A.J.T.

submitted to the EMBL Data Library, January 1995

A:Description: Identification of a 38kDa heparin-binding glycoprotein (gp38k) in differ

A:Reference number: S51327

A:Accession: S51327

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-383 &lt;SHA&gt;

A:Cross-references: EMBL:Z47803; NID:G634097; PIDN:CAA87764.1; PID:G634098

C:Superfamily: Streptomyces chitinase chl40

Query Match 60.2%; Score 50; DB 2; Length 383;  
Best Local Similarity 50.0%; Pred. No. 0.5;

Matches 8; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 1 DKLVGVPFYGRAXSI 16  
:|||||: :|||:  
Db 252 NKLVGIPFGRSFTL 267

## RESULT 15

JC4565

chitinase (EC 3.2.1.14) 1 precursor - *Coccidioides immitis*  
N:Alternate names: complement fixation antigen homolog; CT51 protein  
C:Species: *Coccidioides immitis*  
C:Date: 12-Mar-1996 #sequence\_revision 19-Apr-1996 #text\_change 13-Nov-1998

C:Accession: JC4565

R:Pishko, E.U.; Kirkland, T.N.; Cole, G.T.

Gene 167; 173-177; 1995

A:Title: Isolation and characterization of two chitinase-encoding genes (ctsl, cts2) from

A:Reference number: JC4565; MUID:96144270; PMID:8566773

A:Accession: JC4565

A:Molecule type: mRNA

A:Residues: 1-427 &lt;PIS&gt;

A:Cross-references: GB:U41663

A:Experimental source: C735

C:Genetics:

A:Gene: cts1

A:introns: 47/3; 171/3; 191/3; 215/2; 393/3

C:Superfamily: Streptomyces chitinase chl40

C:Keywords: glycoprotein; glycosidase; hydrolase; polysaccharide degradation

F:1-17/Domain: signal sequence #status predicted &lt;Sig&gt;

F:18-427/Product: chitinase 1 #status predicted &lt;MAT&gt;

F:387/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 60.2%; Score 50; DB 2; Length 427;  
Best Local Similarity 61.5%; Pred. No. 0.56;  
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKLVGVPFYGRA 13  
:|||||:  
Db 284 NKVLGMPFYGRA 296

Search completed: March 22, 2004, 07:01:26  
Job time: 1.10561 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:31:13 ; Search time 0.652029 Seconds

(without alignments)  
1357.597 Million cell updates/sec

Title: US-09-662-293-10

Perfect score: 83  
Sequence: 1 DKLVGVFYGKXSTIE 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	61	73.5	499	1	CHIB_SERMA
2	58	69.9	554	1	CHIT_MANSE
3	56	67.5	458	1	CHIT_DROME
4	56	67.5	699	1	CHIT_BACCI
5	54	65.1	508	1	CHIT_DROME
6	51	61.4	623	1	OGP_BAPAN
7	51	61.4	678	1	OGP_HUMAN
8	51	61.4	721	1	OGP_MOUSE
9	50	60.2	427	1	CHIT_COEPO
10	49	59.0	383	1	CHIT_HUMAN
11	49	59.0	385	1	DIAC_HUMAN
12	49	59.0	671	1	OGP_MESAU
13	48	57.8	367	1	DIAC_RAT
14	48	57.8	610	1	CHIT_STRPL
15	48	57.8	619	1	CHIT_STRPL
16	47	56.6	390	1	CHIT_HUMAN
17	47	56.6	396	1	CHIT_MOUSE
18	47	56.6	423	1	CHIT_APHAL
19	47	56.6	423	1	CHIT_TRIHA
20	47	56.6	466	1	CHIT_HUMAN
21	47	56.6	537	1	OGP_BOVIN
22	47	56.6	539	1	OGP_SHEEP
23	45	54.2	527	1	OGP_PIG
24	45	54.2	820	1	CHIT_ALISO
25	44	53.0	617	1	CHIT_CAEBL
26	43	51.8	185	1	NISG_IACIA
27	43	51.8	381	1	CHIT_MOUSE
28	42	50.6	3010	1	POLG_HCVK
29	42	50.6	3010	1	POLG_HCVK
30	42	50.6	3010	1	POLG_HCVK
31	41	49.8	1226	1	PAIT_CAEBL
32	40.5	48.4	2376	1	TA03_YEAST
33	40	48.2	207	1	ACPD_ANASP

34	40	48.2	504	1	CHIT_BRUMA
35	40	48.2	550	1	CHIT_NPYOP
36	40	48.2	3010	1	POLG_HCVT
37	40	48.2	3011	1	POLG_HCVL
38	40	48.2	3011	1	POLG_HCVL
39	40	48.2	3122	1	DPOZ_MOUSE
40	40	48.2	3130	1	DPOZ_HUMAN
41	39	47.0	551	1	CHIT_NPVAC
42	39	47.0	555	1	LUCI_VARI
43	39	47.0	1504	1	DPOZ_YEAST
44	38	45.8	476	1	CHIT_HUMAN
45	38	45.8	1314	1	TERX_CLOTE

## ALIGNMENTS

RESULT 1	ID	CHIB_SERMA	STANDARD;	PRT;	499 AA.
AC	PI1797	CHIB_SERMA			
DT	01-OCT-1989	(Rel. 12, Created)			
DT	01-OCT-1989	(Rel. 12, Last sequence update)			
DT	30-MAY-2000	(Rel. 39, Last annotation update)			
DE	Chitinase B precursor (EC 3.2.1.14).				
GN	CHIB				
OS	Serratia marcescens.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
OC	Enterobacteriaceae; Serratia.				
OX	NCBI_TaxId=615;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRATN=2170;				
RC	MEDLINE=98037511; PubMed=9371460;				
RA	Watanabe T., Kimura K., Sumiya T., Nikaidou N., Suzuki K., Suzuki M.,				
RA	Taiyogi M., Ferrer S., Regue M.;				
RT	Genetic analysis of the chitinase system of Serratia marcescens				
RT	2170. "				
RL	J. Bacteriol. 179:7111-7117(1997).				
CC	-1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-				
CC	acetyl-D-glucosamine polymers of chitin.				
CC	-1- SIMILARITY: Belongs to chitinase class II (family 18 of glycosyl				
CC	hydrolases).				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
DR	EMBL; X15208; CAA3278.1; -				P29030 brugia mala
DR	EMBL; AB015997; BAA31568.1; -				O10363 oryza pseu
DR	PIR; S04856; S04856.				000269 h genome po
DR	HSSP; P07254; ICTN.				P26664 h genome po
DR	InterPro; IPR003610; CBM_5_12.				P27958 h genome po
DR	InterPro; IPR001223; Glyco_hydro_18.				O61493 mus musculu
DR	InterPro; IPR001579; Glyco_hydro_18as.				O60673 homo sapien
DR	Pfam; PF02839; CBM_5_12; 1.				P41684 autographa
DR	Pfam; PF00704; Glyco_hydro_18; 1.				P17554 vargula hll
DR	ProDom; PD000471; Glyco_hydro_18; 1.				P14284 saccharomyc
DR	SMART; SM00495; ChnB3; 1.				O9b2p6 homo sapien
DR	SMART; SM00636; Glyco_18; 1.				P04958 clostridium
DR	PROSITE; PS01095; CHITINASE_18; 1.				

KW Hydrolase; Glycosidase; Chitin degradation; Signal.  
 FT SIGNAL 1 41  
 FT CHAIN 42 499  
 FT ACT SITE 144 144  
 DR PROSITE; PSS0940; CHIT\_BIND\_II; 1.  
 DR POSITIVE; PSS01095; CHITINASE\_18; 1.  
 KW Hydrolase; Glycosidase; Chitin degradation; Chitin-binding; Signal;  
 SQ SEQUENCE 499 AA; 55464 MW; FFD674916109D1D8 CRC64;  
 Query Match 73.5%; Score 61; DB 1; Length 499;  
 Best Local Similarity 91.7%; Pred. No. 0.0041;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Db 2 KLMGVGFYGRA 13  
 264 KLMGVGFYGRA 295  
 RESULT 2  
 CHIT\_MANSF STANDARD; PRT; 554 AA.  
 ID CHIT\_MANSF STANDARD; PRT; 554 AA.  
 AC P36362;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Endochitinase precursor (EC 3.2.1.14).  
 OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Sphingioidea;  
 OC Sphingidae; Sphinginae; Manduca.  
 OX NCBI\_TaxID=7130;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=93537793; PubMed=8353525;  
 RA Kramer K.J., Corpuz L., Choi H.K., Muthukrishnan S.;  
 RT "Sequence of a cDNA and expression of the gene encoding epidermal and  
 RT gut chitinases of *Manduca sexta*."  
 RT Insect Biochem. Mol. Biol. 23:691-701(1993).  
 RL Insect Biochem. Mol. Biol. 23:691-701(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=97215580; PubMed=9061927;  
 RA Choi H.K., Choi K.H., Kramer K.J., Muthukrishnan S.;  
 RT "Isolation and characterization of a genomic clone for the gene of an  
 RT insect molting enzyme, chitinase."  
 RT Insect Biochem. Mol. Biol. 27:37-47(1997).  
 RL Insect Biochem. Mol. Biol. 27:37-47(1997).  
 RN [3]  
 RP FUNCTION: Digest chitin in the exoskeleton during the molting  
 process.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-  
 CC acetyl-D-glucosamine polymers of chitin.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: Epidermis and gut.  
 CC -1- DEVELOPMENTAL STAGE: High levels seen in the epidermis on day 0,  
 CC but rapidly disappears and is undetected on days 1-4 of fifth  
 CC instar. It reappears on day 5 and peaks on day 7 after which a  
 CC levels decline is seen. In the gut is detected on day 6 with lower  
 CC levels seen on days 0, 7 and 8.  
 CC -1- SIMILARITY: Belongs to chitinase class II (family 18 of glycosyl  
 CC hydrolases).  
 CC -1- SIMILARITY: Contains 1 chitin-binding type-2 domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; U02270; AAC04924.1;  
 CC DR EMBL; L49234; AAB53952.1;  
 CC PIR; A56596; A56596.  
 CC InterPro; IPR002557; Chitin\_bind\_Pera.  
 CC InterPro; IPR001223; Glyco\_hydro\_18.  
 CC InterPro; IPR001579; Glyco\_hydro\_18AS.  
 CC Pfam; PF01607; CBM\_14; 1.  
 CC Pfam; PF00704; Glyco\_hydro\_18; 1.

DR Prodom; PD000471; Glyco\_hydro\_18; 1.  
 DR SMART; SM00494; ChitBD; 1.  
 DR SMART; SM00636; Glyco\_18; 1.  
 DR PROSITE; PSS0940; CHIT\_BIND\_II; 1.  
 DR POSITIVE; PSS01095; CHITINASE\_18; 1.  
 KW Hydrolase; Glycosidase; Chitin degradation; Chitin-binding; Signal;  
 SQ SEQUENCE 554 AA; 62203 MW; 3989D756C96CD490 CRC64;  
 Query Match 69.3%; Score 58; DB 1; Length 554;  
 Best Local Similarity 56.2%; Pred. No. 0.015;  
 Matches 9; Conservative 6; Mismatches 1; Indels 0; Gaps 0;  
 Db 1 DKLWGVGFYGRAXSI 16  
 261 NKLVGVGFYGRSFTL 276  
 RESULT 3  
 CHIT\_DROME STANDARD; PRT; 458 AA.  
 ID CHIT\_DROME STANDARD; PRT; 458 AA.  
 AC Q9W5U2; O17422;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Probable chitinase 3 (EC 3.2.1.14).  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=Berkely;  
 RA MEDLINE=22426071; PubMed=12537574;  
 RA Hoskins R.A., Smith C.D., Carlson J.W., Carvalho A.B., Halpern A.,  
 RA Kaminber J.S., Kennedy C., Mungall C.D., Sullivan B.A., Sutton G.G.,  
 RA Yasuhara J.C., Wakimoto B.T., Myers E.W., Ceiniker S.E., Rubin G.M.,  
 RA Karpen G.H.;  
 RT "Heterochromatic sequences in a *Drosophila* whole-genome shotgun  
 RT assembly."  
 RT Genome Biol. 3:RESEARCH0085.1-RESEARCH0085.16(2002).  
 RL Genome Biol. 3:RESEARCH0085.1-RESEARCH0085.16(2002).  
 RN [2]  
 RP SEQUENCE OF 182-294 FROM N.A.  
 RA STRAIN=Canton-S;  
 RA MEDLINE=88324849; PubMed=9662472;  
 RA de la Vega H., Specht C.A., Liu Y., Robbins P.W.;  
 RT "Chitinases are a multi-gene family in *Aedes*, *Anopheles* and  
 RT *Drosophila*."  
 RT Insect Mol. Biol. 7:233-239(1998).  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-  
 CC acetyl-D-glucosamine polymers of chitin.  
 CC -1- SIMILARITY: Belongs to chitinase class II (family 18 of glycosyl  
 CC hydrolases).  
 CC -1- SIMILARITY: Contains 2 chitin-binding type-2 domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC -----
DR EMBL: AF026502; AAB81860.1; -.
DR FlyBase: FBgn022701; Chit3.
DR InterPro: IPR002557; Chitin_bind_Pera.
DR InterPro: IPR001223; Glyco_hydro_18.
DR InterPro: IPR01579; Glyco_hydro_18AS.
DR Pfam: PF01607; CBM_14; 2.
DR Pfam: PF00704; Glyco_hydro_18; 1.
DR ProDom: PD000471; Glyco_hydro_18; 1.
DR SMART: SM00494; ChitBD2; 2.
DR SMART: SM00636; Glyco_18; 1.
DR PROSITE: PS00940; CHIT_BIND_II; 2.
DR PROSITE: PS01095; CHITINASE_18; FALSE NEG.
KM Hydrolyase; Glycosidase; Chitin degradation; Chitin-binding;
KM Multigene family; Repeat.
FT DOMAIN 5 58 CHITIN-BINDING TYPE-2 1.
FT ACT_SITE 74 128 CHITIN-BINDING TYPE-2 2.
FT ACT_SITE 295 295 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 458 AA; 52330 MW; 4A063190B7B96248 CRC64;

Query Match 67.5%; Score 56; DB 1; Length 458;
Best Local Similarity 56.2%; Pred. No. 0.028; 2; Indels 0; Gaps 0;
Matches 9; Conservative 5; Mismatches 2;

Oy 2 KLVMGVPEYGRAXSIE 17
Db 410 KLVMGIFLYGQSFTLE 425

RESULT 4
CHIT_BACCI STANDARD; PRT; 699 AA.
ID CHIT_BACCI STANDARD; PRT; 699 AA.
AC P20533;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Chitinase A1 precursor (EC 3.2.1.14).
GN CHIT1.
OS Bacillus circulans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_Taxid=1397;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WL-12;
RX MEDLINE=9036876; PubMed=2203782;
RA Watanabe T., Suzuki K., Oyangi W., Ohnishi K., Tanaka H.;
RT "Gene cloning of chitinase A1 from Bacillus circulans WL-12 revealed its evolutionary relationship to Serratia chitinase and to the type III homology units of fibronectin."
RL J. Biol. Chem. 265:15659-15665(1990).
RN [2]
RP MUTAGENESIS.
RC STRAIN=WL-12;
RX MEDLINE=93365760; PubMed=8103047;
RA Watanabe T., Kohori K., Miyashita K., Fujii T., Sakai H.,
RA Uchida M., Tanaka H.;
RT "Identification of glutamic acid 204 and aspartic acid 200 in chitinase A1 of Bacillus circulans WL-12 as essential residues for chitinase activity."
RL J. Biol. Chem. 268:18567-18572(1993).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-acetyl-D-glucosamine polymers of chitin.
CC -1- SIMILARITY: Contains 2 fibronectin type III domains.
CC -1- SIMILARITY: Belongs to chitinase class II (family 18 of glycosyl hydrolases).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ebi.ac.uk/announcements or send an email to license@ebi.ac.uk).

```

```

CC -----
DR EMBL: M57601; AAB81828.1; -.
DR PIR: A38368; A38368.
DR PDB: 1ED7; 24-MAY-00.
DR PDB: 1K65; 18-DEC-02.
DR InterPro: IPR003610; CBM_5_12.
DR InterPro: IPR008957; FN_III-like.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR001223; Glyco_hydro_18.
DR InterPro: IPR01579; Glyco_hydro_18AS.
DR Pfam: PF02839; CBM_5_12; 1.
DR Pfam: PF00411; fn3; 2.
DR Pfam: PF00704; Glyco_hydro_18; 1.
DR ProDom: PD000471; Glyco_hydro_18; 2.
DR SMART: SM00495; ChitBD3; 1.
DR SMART: SM00600; FN3; 2.
DR SMART: SM00636; Glyco_18; 1.
DR PROSITE: PS01095; CHITINASE_18; 1.
KM Hydrolyase; Glycosidase; Chitin degradation; Signal; Repeat;
KM 3D-structure.
FT SIGNAL 1 41
FT CHAIN 42 699 CHITINASE A1.
FT DOMAIN 42 460 CATALYTIC.
FT DOMAIN 465 549 FIBRONECTIN TYPE-III (R-1).
FT DOMAIN 560 644 FIBRONECTIN TYPE-III (R-2).
FT ACT_SITE 204 204 PROTON DONOR (PROBABLE).
FT MUTAGEN 200 200 D->N: DECREASE IN ACTIVITY.
FT MUTAGEN 200 200 D->E: NO CHANGE IN ACTIVITY.
FT MUTAGEN 204 204 E->D/Q: LOSS OF ACTIVITY.
SQ SEQUENCE 699 AA; 73677 MW; AC7C9B2E2987643 CRC64;

Query Match 67.5%; Score 56; DB 1; Length 699;
Best Local Similarity 90.9%; Pred. No. 0.044; 1; Indels 0; Gaps 0;
Matches 10; Conservative 1; Mismatches 0;

Oy 2 KLVMGVPEYGR 12
Db 330 KLVMGIFLYGR 340

RESULT 5
CHIT_DROME STANDARD; PRT; 508 AA.
ID CHIT_DROME STANDARD; PRT; 508 AA.
AC Q9M5U3; O17420;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable chitinase 1 (EC 3.2.1.14).
GN CHIT1 OR CG17682.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=22426071; PubMed=12537574;
RA Hoskins R.A., Smith C.D., Carlson J.W., Carvalho A.B., Halpern A.,
RA Kaminker J.S., Kennedy C., Mungall C.J., Sullivan B.A., Sutton G.G.,
RA Yasunara J.C., Wakimoto B.T., Myers E.W., Celniker S.E., Rubin G.M.,
RA Karpen G.H.;
RT "Heterochromatic sequences in a Drosophila whole-genome shotgun assembly."
RL Genome Biol. 3:RESEARCH0085.1-RESEARCH0085.16(2002).
RN [2]
RP SEQUENCE OF 151-263 FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=98324849; PubMed=9662472;
RA de la Vega H., Specht C.A., Liu Y., Robbins P.W.;
RT "Chitinsases are a multi-gene family in Aedes, Anopheles and Drosophila."
RL Insect Mol. Biol. 7:233-239(1998).

```

```

CC -----
CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-
CC acetyl-D-glucosamine polymers of chitin.
CC -1- SIMILARITY: Belongs to chitinase class II (family 18 of glycosyl
CC hydrolases).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF026500; AAB81858.1; -
CC FLYBASE; FBgn0022703; Chit1.
CC InterPro; IPR001223; Glyco_hydro_18.
CC InterPro; IPR001579; Glyco_hydro_18AS.
CC Pfam; PF00704; Glyco_hydro_18; 2.
CC ProDom; PD000471; Glyco_hydro_18; 2.
CC SMART; SM00636; Glyco_18; 1.
CC DR PROSITE; PS01095; CHITINASE_18; 1.
CC DR PROSITE; PS01095; CHITINASE_18; 2.
CC KEGG; HYDROLASE; Glycosidase; Chitin degradation; Multisene family.
CC ACT_SITE 264 264 PROTON DONOR (BY SIMILARITY).
CC SEQUENCE 308 AA; 57751 MW; 26CA23B02FEDE97 CRC64;
SQ
Query Match 65.1%; Score 54; DB 1; Length 508;
Best Local Similarity 60.0%; Pred. No. 0.071;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
QY 2 KLVMGVFPYGRAXSI 16
Db 378 KLVMGIPLYGQSFSL 392

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M59903; AAB39765.1; -
CC PIR; A37954; A37954.
CC InterPro; IPR001223; Glyco_hydro_18.
CC InterPro; IPR001579; Glyco_hydro_18AS.
CC Pfam; PF00704; Glyco_hydro_18; 1.
CC ProDom; PD000471; Glyco_hydro_18; 1.
CC SMART; SM00636; Glyco_18; 1.
CC DR PROSITE; PS01095; CHITINASE_18; FALSE_NEG.
CC KEGG; GLYCOPROTEIN; Fertilization; Signal.
CC FT SIGNAL 1 21 BY SIMILARITY.
CC FT CHAIN 22 623 OVIDUCT-SPECIFIC GLYCOPROTEIN.
CC FT CARBOHYD 402 402 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 441 441 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 623 AA; 69291 MW; 9E21CF481PFF1268 CRC64;
SQ
Query Match 61.4%; Score 51; DB 1; Length 623;
Best Local Similarity 66.7%; Pred. No. 0.3;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 DKLVMGVFPYGR 12
Db 254 EKLIMGIPYGR 265

```

```

RESULT 6
OGP_PAPAN STANDARD; PRT; 623 AA.
ID ID_OGP_PAPAN STANDARD; PRT; 623 AA.
AC P36718;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Oviduct-specific glycoprotein precursor (Oviductal glycoprotein)
DE (Oviductin) (Estrogen-dependent oviduct protein).
GN OVGPI OR OGP.
OS Papio anubis (Olive baboon).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Papio.
OX NCBI_TaxID=9555;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Oviduct;
RX MEDLINE=98244335; PubMed=9584944;
RA Verhage H.G., Fazleabas A.T., Mavrogianis P.A., O'Day-Bowman M.B.,
RA Donnelly K.M., Arias E.B., Jaffe R.C.;
RT "The baboon oviduct: characteristics of an oestradiol-dependent
RT oviduct-specific glycoprotein."
RL Hum. Reprod. Update 3:541-552(1997).
RN [2]
RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RC TISSUE=Oviduct;
RX MEDLINE=91367180; PubMed=1716345;
RA Jaffe R.C.;
RA Donnelly K.M., Fazleabas A.T., Verhage H.G., Mavrogianis P.A.,
RT "Cloning of a recombinant complementary DNA to a baboon (Papio
RT anubis) estradiol-dependent oviduct-specific glycoprotein."
RL Mol. Endocrinol. 5:356-364(1991).
RN [3]
RP FUNCTION: Binds to oocyte zona pellucida in vivo. May play a role
RP in the fertilization process and/or early embryonic development.
CC -1- SUBCELLULAR LOCATION: Secretory granules.
CC -1- TISSUE SPECIFICITY: Oviduct.
CC -1- DEVELOPMENTAL STAGE: At the time of ovulation.
CC -1- SIMILARITY: Belongs to family 18 of glycosyl hydrolases.

```

```

RESULT 7
OGP_HUMAN STANDARD; PRT; 678 AA.
ID ID_OGP_HUMAN STANDARD; PRT; 678 AA.
AC Q12889; Q15841;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Oviduct-specific glycoprotein precursor (Oviductal glycoprotein)
DE (Oviductin) (Estrogen-dependent oviduct protein) (Mucin 9).
GN OVGPI OR OGP OR MUC9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Oviduct;
RX MEDLINE=95119256; PubMed=7819450;
RA Arias E.B., Verhage H.G., Jaffe R.C.;
RT "Complementary deoxyribonucleic acid cloning and molecular
RT characterization of an estrogen-dependent human oviductal
RT glycoprotein."
RL Biol. Reprod. 51:685-694(1994).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANT GLN-676.
RC Jaffe R.C.;
RA Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC Jaffe R.C.;
RA Coville G.;
RT Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Binds to oocyte zona pellucida in vivo. May play a role
CC in the fertilization process and/or early embryonic development.
CC -1- SUBCELLULAR LOCATION: Secretory granules.
CC -1- TISSUE SPECIFICITY: Oviduct.
CC -1- SIMILARITY: Belongs to family 18 of glycosyl hydrolases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
DR EMBL; U09550; AAA86946.1; -  
DR EMBL; U58010; AAB04126.1; -  
DR EMBL; U58001; AAB04126.1; JOINED.  
DR EMBL; U58002; AAB04126.1; JOINED.  
DR EMBL; U58003; AAB04126.1; JOINED.  
DR EMBL; U58004; AAB04126.1; JOINED.  
DR EMBL; U58005; AAB04126.1; JOINED.  
DR EMBL; U58006; AAB04126.1; JOINED.  
DR EMBL; U58007; AAB04126.1; JOINED.  
DR EMBL; U58008; AAB04126.1; JOINED.  
DR EMBL; U58009; AAB04126.1; JOINED.  
DR EMBL; A1390195; CAC36039.1; -  
DR Genew; H9NC; 8524; OVGP1.  
DR MIM; 603578; -  
DR GO; GO:0007565; P:pregnancy; TAS.  
DR Interpro; IPR001223; Glyco\_hydro\_18.  
DR Pfam; PF00704; Glyco\_hydro\_18S.  
DR ProDom; PD000471; Glyco\_hydro\_18; 1.  
DR SMART; SM00636; Glyco\_18; 1.  
DR PROSITE; PS01095; CHITINASE\_18; FALSE NEG.  
KW Glycoprotein; Fertilization; Signal; Polymorphism.  
FT SIGNAL 1 21  
FT CHAIN 22 678  
FT CARBOHYD 402 402  
FT CARBOHYD 441 441  
FT CARBOHYD 580 580  
FT CARBOHYD 596 596  
FT CARBOHYD 648 648  
FT VARIANT 676 676  
FT CONFLICT 477 477  
FT CONFLICT 511 511  
FT CONFLICT 514 514  
SQ SEQUENCE 678 AA; 75421 MW; 245F2CEDCE92768B CRC64;

Query Match 61.4%; Score 51; DB 1; Length 678;  
Best Local Similarity 66.7%; Pred. No. 0.32;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKLWGVFPYGR 12  
Db 254 EKLINGIPTYGR 265

-----  
RESULT 8  
OGP MOUSE  
ID OGP MOUSE STANDARD; PRT; 721 AA.  
AC 062010;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Oviduct-specific glycoprotein precursor (Oviductal glycoprotein)  
DE (Oviductin) (Barrigen-dependent oviduct protein).  
GN OVGP1 OR OGP OR CHITS.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ICR; TISSUE=Oviduct;  
RX MEDLINE=96115001; PubMed=7492680;  
RA Sendai Y., Komiya H., Suzuki K., Onuma T., Kikuchi M., Hoshi H.,  
RA Araki Y.;  
RT "Molecular cloning and characterization of a mouse oviduct-specific  
RT glycoprotein."  
RL Biol. Reprod. 53:285-294 (1995).

CC - FUNCTION: Binds to oocyte zona pellucida in vivo. May play a role  
CC in the fertilization process and/or early embryonic development.  
CC - SUBCELLULAR LOCATION: Secretory granules.  
CC - TISSUE SPECIFICITY: Epithelial cells of the oviduct.  
CC - SIMILARITY: Belongs to family 18 of glycosyl hydrolases.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
DR EMBL; D32137; BAA06863.1; -  
DR MGD; MGI:106661; OVGP1.  
DR Interpro; IPR001223; Glyco\_hydro\_18.  
DR Interpro; IPR001579; Glyco\_hydro\_18S.  
DR Pfam; PF00704; Glyco\_hydro\_18; 1.  
DR ProDom; PD000471; Glyco\_hydro\_18; 1.  
DR SMART; SM00636; Glyco\_18; 1.  
DR PROSITE; PS01095; CHITINASE\_18; FALSE NEG.  
KW Glycoprotein; Fertilization; Repeat; Signal.  
FT SIGNAL 1 21  
FT CHAIN 22 721  
FT DOMAIN 486 632  
FT CARBOHYD 402 402  
FT CARBOHYD 442 442  
FT CARBOHYD 469 469  
SQ SEQUENCE 721 AA; 78807 MW; 37246C8F0165652 CRC64;

Query Match 61.4%; Score 51; DB 1; Length 721;  
Best Local Similarity 75.0%; Pred. No. 0.34;  
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 DKLWGVFPYGR 12  
Db 254 EKLINGIPTYGR 265

-----  
RESULT 9  
CHIT1\_COCPO  
ID CHIT1\_COCPO STANDARD; PRT; 427 AA.  
AC P54196;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Endochitinase 1 precursor (EC 3.2.1.14) (Complement-fixation antigen)  
DE (CF-antigen) (CF-Ag).  
GN CTS1.  
OS Coccidioides posadasii.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Onygenales; mitosporic Onygenales; Coccidioides.  
OX NCBI\_TaxID=199306;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C735;  
RX MEDLINE=96144270; PubMed=8566773;  
RA Piskho E.J., Kirkland T.N., Cole G.T.;  
RT Isolation and characterization of two chitinase-encoding genes  
RT (cts1, cts2) from the fungus Coccidioides immitis."  
RL Gene 167:173-177 (1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Silveira;  
RA Yang C., Zhu Y., Magee D.M., Cox R.A.;  
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
CC - CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-  
CC acetyl-D-glucosamine polymers of chitin.  
CC - SIMILARITY: Belongs to chitinase class II (family 18 of glycosyl  
CC hydrolases).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; L41663; AAA92643.1; -;  
CC EMBL; U51271; AAA96515.1; -;  
CC EMBL; U33265; AAB06687.1; -;  
CC PDB; 1D2K; 27-SEP-00.  
CC PDB; 1LL4; 25-SEP-02.  
CC PDB; 1LL6; 18-DEC-02.  
CC PDB; 1LL7; 18-DEC-02.  
CC InterPro; IPR01223; Glyco\_hydro\_18.  
CC InterPro; IPR01579; Glyco\_hydro\_18A5.  
CC Pfam; PF00704; Glyco\_hydro\_18; 1.  
CC ProDom; PD000471; Glyco\_hydro\_18; 1.  
CC SMART; SM00636; Glyco\_18; 1.  
CC PROSITE; PS01095; CHITINASE\_18; 1.  
CC Hydroclase; Glycosidase; Chitin degradation; Chitin-binding; Signal;  
CC Glycoprotein; 3D-structure.  
CC FT CHAIN 1 38 POTENTIAL.  
FT CARBOHYD 387 387 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CONFLICT 15 47 RWLSRLCLFELGFMFTLSTAVLVTDIQ -> VQAS  
FT CONFLICT 15 47 SMSMSNPYVPYEPABEGFSPVYFVM (IN REF. 2).  
FT CONFLICT 199 199 K -> N (IN REF. 2).  
SQ SEQUENCE 427 AA; 47629 MW; 1C396DBD1A7001A CRC64;

Query Match 60.2%; Score 50; DB 1; Length 427;  
Best Local Similarity 61.5%; Pred. No. 0.3;  
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKLVGVPFYGRA 13  
Db 284 NKIVLGMPLYGRA 296

RESULT 10  
C3L1 HUMAN STANDARD; PRT; 383 AA.  
AC P36222; P30923;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Chitinase-3 like protein 1 precursor (Cartilage glycoprotein-39) (GP-39) (39 kDa synovial protein) (YKL-40).  
DE CH3L1.  
GN Homo sapiens (Human).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
RC TISSUE=Cartilage;  
RX MEDLINE=94064659; PubMed=8245017;  
RA Hakala B.E., White C., Recklies A.D.;  
RT "Human cartilage gp-39, a major secretory product of articular chondrocytes and synovial cells, is a mammalian member of a chitinase protein family."  
RL T. Biol. Chem. 268:25803-25810(1993).  
[2]  
RN SEQUENCE FROM N.A.  
RC TISSUE=Blood;  
RX MEDLINE=97386591; PubMed=9244440;  
RA Reili M., Krause S.W., Andresen R.;  
RT "Molecular characterization of the gene for human cartilage gp-39 (CH3L1), a member of the chitinase protein family and marker for late stages of macrophage differentiation."  
RL Genomics 43:221-225(1997).  
[3]

RP SEQUENCE OF 22-45.

RX MEDLINE=90328983; PubMed=2375755;  
RA Nytko P., Golds B.E.;  
RT "Human synovial cells secrete a 39 kDa protein similar to a bovine mammary protein expressed during the non-lactating period."  
RL Biochem. J. 269:265-268(1990).  
CC -1- FUNCTION: May play an important role in the capacity of cells to respond to and cope with changes in their environment.  
CC -1- SUBCELLULAR LOCATION: Extracellular.  
CC -1- TISSUE SPECIFICITY: Present in articular chondrocytes, synovial cells as well as in liver. Undetectable in muscle tissues, lung, pancreas, mononuclear cells, or fibroblasts.  
CC -1- PTM: Glycosylated.  
CC -1- SIMILARITY: Belongs to family 18 of glycosyl hydrolases.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M80927; AAA16074.1; -;  
CC EMBL; Y08374; CAA69661.1; JOINED.  
CC EMBL; Y08375; CAA69661.1; JOINED.  
CC EMBL; Y08376; CAA69661.1; JOINED.  
CC EMBL; Y08377; CAA69661.1; JOINED.  
CC EMBL; Y08378; CAA69661.1; JOINED.  
CC PIR; A49562; A49562.  
CC PDB; 1LA7; 10-APR-02.  
CC Genew; HGNC:1932; CH3L1.  
CC MIM; 601525; -;  
CC DR GO; GO:0005788; Extracellular matrix; TAS.  
CC DR GO; GO:0005615; Extracellular space; TAS.  
CC DR GO; GO:0005201; Extracellular matrix structural constituent; TAS.  
CC InterPro; IPR01223; Glyco\_hydro\_18.  
CC InterPro; IPR01579; Glyco\_hydro\_18A5.  
CC Pfam; PF00704; Glyco\_hydro\_18; 1.  
CC ProDom; PD000471; Glyco\_hydro\_18; 1.  
CC SMART; SM00636; Glyco\_18; 1.  
CC PROSITE; PS01095; CHITINASE\_18; FALSE\_NEG.  
CC Glycoprotein; Signal; 3D-structure.  
CC FT CHAIN 1 21 CHITINASE-3 LIKE PROTEIN 1.  
FT CARBOHYD 60 60 N-LINKED (GLCNAC... ) (POTENTIAL).  
SQ SEQUENCE 383 AA; 42613 MW; 76ADD8298BEEC2D1 CRC64;

Query Match 59.0%; Score 49; DB 1; Length 383;  
Best Local Similarity 53.3%; Pred. No. 0.41;  
Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 KLVGVPFYGRA 16  
Db 253 KLVGVPFYGRA 267

RESULT 11  
DIAC HUMAN STANDARD; PRT; 385 AA.  
AC Q01459;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE D1-N-acetylchitinase precursor (EC 3.2.1.-).  
GN CTBS OR CTB.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RN SEQUENCE FROM N.A.  
RC TISSUE=Placenta;

RA MEDLINE=92406917; PubMed=1527079;  
 RA Fisher K.J.; Aronson N.N. Jr.;  
 RT "Cloning and expression of the CDNA sequence encoding the lysosomal  
 RT glycosylase di-N-acetylchitobiose.";  
 RL J. Biol. Chem. 267:19607-19616(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9270791; PubMed=10336991;  
 RA Liu B., Ahmad W., Aronson N.N. Jr.;  
 RT "Structure of the human gene for lysosomal di-N-acetylchitobiose.";  
 RL Glycobiology 9:589-593(1999).  
 CC -1- FUNCTION: Involved in the degradation of asparagine-linked  
 CC glycoproteins. Hydrolyze of N-acetyl-beta-D-glucosamine  
 CC (1-4)-N-acetylglucosamine chitobiose core from the reducing end  
 CC of the bond, it requires prior cleavage by glycosylasparaginase.  
 CC -1- SUBCELLULAR LOCATION: Lysosomal.  
 CC -1- SIMILARITY: Belongs to family 18 of glycosyl hydrolases.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M95767; AAC35684.1; -.  
 DR EMBL: AF085706; AAC35852.1; -.  
 DR EMBL: AF085700; AAC35852.1; JOINED.  
 DR EMBL: AF085701; AAC35852.1; JOINED.  
 DR EMBL: AF085702; AAC35852.1; JOINED.  
 DR EMBL: AF085703; AAC35852.1; JOINED.  
 DR EMBL: AF085704; AAC35852.1; JOINED.  
 DR EMBL: AF085705; AAC35852.1; JOINED.  
 DR PIR: A44102; A44102.  
 DR Genew: HGNC:2496; CTBS.  
 DR MIM: 600873; -.  
 DR InterPro: IPR001223; Glyco\_hydro\_18.  
 DR InterPro: IPR001579; Glyco\_hydro\_18s.  
 DR Pfam: PF00704; Glyco\_hydro\_18; 1.  
 DR Prodom: PD000471; Glyco\_hydro\_18; 1.  
 DR SMART: SM00636; Glyco\_18; 1.  
 DR PROSITE: PS01095; CHITINASE\_18; 1.  
 KW Hydrolase; Glycosidase; Signal; Lysosome; Glycoprotein.  
 FT SIGNAL 1 38  
 FT CHAIN 39 385  
 FT ACT\_SITE 143 143 DI-N-ACETYLCHITOBIOSE.  
 FT CARBOHYD 193 193 PROTON DONOR (BY SIMILARITY).  
 FT CARBOHYD 228 228 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 385 AA; 43759 MW; 0A9D14CB82B52BE CRC64;  
 Query Match 59.0%; Score 49; DB 1; Length 385;  
 Best Local Similarity 90.0%; Pred. NO. 0.41;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 CC Mesocricetus.  
 CC NCBI\_TaxID=10036;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE-Oviduct;  
 RX MEDLINE=96115007; PubMed=7492686;  
 RA Suzuki K., Sendai Y., Onuma T., Hoshi H., Hiroi M., Araki Y.;  
 RT "Molecular characterization of a hamster oviduct-specific  
 RT glycoprotein.";  
 RL Biol. Reprod. 53:345-354(1995).  
 RN [2]  
 RP SEQUENCE OF 22-671 FROM N.A.  
 RC TISSUE-Oviduct;  
 RX MEDLINE=96192955; PubMed=8607967;  
 RA Paquette Y., Merlen Y., Malette B., Bleau G.;  
 RT "Allelic polymorphism in the hamster oviductin gene is due to a  
 RT variable number of mucin-like tandem repeats.";  
 RL Mol. Reprod. Dev. 42:388-396(1995).  
 RN [3]  
 RP SEQUENCE OF 14-671 FROM N.A., AND REVISIONS.  
 RC TISSUE-Oviduct;  
 RA Paquette Y.;  
 RL Submitted (May-1998) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 22-39.  
 RX MEDLINE=94059981; PubMed=8240241;  
 RA Malette B., Bleau G.;  
 RT "Biochemical characterization of hamster oviductin as a sulphated  
 RT zona pellucida-binding glycoprotein.";  
 RL Biochem. J. 295:437-445(1993).  
 CC -1- FUNCTION: Binds to oocyte zona pellucida in vivo. May play a role  
 CC in the fertilization process and/or early embryonic development.  
 CC Might act as a protective secretion influencing the first steps of  
 CC the reproductive process necessary for the normal triggering of  
 CC fertilization and early embryonic development.  
 CC -1- TISSUE SPECIFICITY: Oviduct.  
 CC -1- SUBCELLULAR LOCATION: Secretory granules.  
 CC -1- PTM: HIGHLY O-GLYCOSYLATED AND ALSO N-GLYCOSYLATED.  
 CC -1- SIMILARITY: Belongs to family 18 of glycosyl hydrolases.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D32218; BA006977.1; -.  
 DR EMBL: U15048; AAC3584.1; -.  
 DR InterPro: IPR001223; Glyco\_hydro\_18.  
 DR InterPro: IPR001579; Glyco\_hydro\_18s.  
 DR Pfam: PF00704; Glyco\_hydro\_18; 1.  
 DR Prodom: PD000471; Glyco\_hydro\_18; 1.  
 DR SMART: SM00636; Glyco\_18; 1.  
 DR PROSITE: PS01095; CHITINASE\_18; FALSE NEG.  
 KW Glycoprotein; Fertilization; Repeat; Signal.  
 FT SIGNAL 1 21  
 FT CHAIN 22 671  
 FT DOMAIN 490 609 OVIDUCT-SPECIFIC GLYCOPROTEIN.  
 FT REPEAT 490 504 8 X 15 AA TANDEN REPEATS.  
 FT REPEAT 505 519 1.  
 FT REPEAT 520 534 2.  
 FT REPEAT 535 549 3.  
 FT REPEAT 550 564 4.  
 FT REPEAT 565 579 5.  
 FT REPEAT 580 594 6.  
 FT REPEAT 595 609 7.  
 FT REPEAT 602 602 8.  
 FT CARBOHYD 402 402 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 511 511 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 526 526 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 526 526



FT CARBOHYD 541 541 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 571 571 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 586 586 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 26 26 C -> A (IN REF. 4).  
 FT CONFLICT 33 33 H -> I (IN REF. 4).  
 FT CONFLICT 137 137 D -> G (IN REF. 2 AND 3).  
 FT CONFLICT 153 153 R -> Q (IN REF. 2 AND 3).  
 FT CONFLICT 165 165 F -> Y (IN REF. 2 AND 3).  
 FT CONFLICT 193 193 Q -> L (IN REF. 2 AND 3).  
 FT CONFLICT 531 545 MISSING (IN REF. 2 AND 3).  
 FT CONFLICT 595 595 T -> I (IN REF. 2 AND 3).  
 SQ SEQUENCE 671 AA; 73250 MW; BB57B0E514EC1972 CRC64;

Query Match 59.0%; Score 49; DB 1; Length 671;  
 Best Local Similarity 75.0%; Pred. No. 0.72;  
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DKLWGVPFYGR 12  
 Db 254 DKLWGVPFYGR 265

## RESULT 13

DIAC\_RAT STANDARD; PRT; 367 AA.  
 AC 001460;  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Di-N-acetylchitobiose precursor (EC 3.2.1.-).  
 GN CTB OR CTB.  
 OS Rattus norvegicus (Rat).  
 OC Buxaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-32.  
 RC TISSUE=Liver;  
 RX MEDLINE=92406917; PubMed=1527079;  
 RA Fisher K.J., Atkinson N.N. Jr.;  
 RT "Cloning and expression of the cDNA sequence encoding the lysosomal  
 glycosylase di-N-acetylchitobiose core from the reducing end of the  
 bond, it requires prior cleavage by glycosylaseparaginase."  
 RU J. Biol. Chem. 267:19607-19616(1992).  
 CC -1- FUNCTION: Involved in the degradation of asparagine-linked  
 glycoproteins. Hydrolyze of N-acetyl-beta-D-glucosamine (1-4)-N-  
 acetylglucosamine chitobiose core from the reducing end of the  
 bond, it requires prior cleavage by glycosylaseparaginase.  
 CC -1- SUBCELLULAR LOCATION: Lysosomal.  
 CC -1- SIMILARITY: Belongs to family 18 of glycosyl hydrolases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M95768; AAA0924.1; -.  
 DR PIR; C44102; C44102.  
 DR InterPro; IPR001223; Glyco\_hydro\_18.  
 DR InterPro; IPR001579; Glyco\_hydro\_18AS.  
 DR Pfam; PF00704; Glyco\_hydro\_18; 1.  
 DR ProDom; PD000471; Glyco\_hydro\_18; 1.  
 DR SMART; SM00636; Glyco\_18; 1.  
 DR PROSITE; PS01095; CHITINASE\_18; 1.  
 KM Hydrolyase; Glycosidase; Signal; Lysosome; Glycoprotein.  
 FT SIGNAL 1 23  
 FT CHAIN 24 367 DI-N-ACETYLCHITOBIOSE.  
 FT ACT\_SITE 128 128 PROTON DONOR (BY SIMILARITY). (POTENTIAL).  
 FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 213 213 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 247 247 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 284 284 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 367 AA; 41531 MW; 29AB8BE4FC157C16 CRC64;

Query Match 57.8%; Score 48; DB 1; Length 367;  
 Best Local Similarity 80.0%; Pred. No. 0.58;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KXLWGVPFYGR 11  
 Db 231 KXLWGVPFYGR 240

## RESULT 14

CHIT\_STRPL STANDARD; PRT; 610 AA.  
 ID CHIT\_STRPL  
 AC P11220;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Chitinase 63 precursor (EC 3.2.1.14).  
 GN CHTA.  
 OS Streptomyces plicatus.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomyces; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1922;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92192480; PubMed=1532161;  
 RA Robbins P.W., Overbye K., Albright C., Benfield B., Pero J.;  
 RT "Cloning and high-level expression of chitinase-encoding gene of  
 Streptomyces plicatus."  
 RU Gene 111:69-76(1992).

CC [2]  
 CC SEQUENCE OF 1-45 FROM N.A., AND SEQUENCE OF 31-45.  
 RX MEDLINE=88087127; PubMed=3275646;  
 RA Robbins P.W., Albright C., Benfield B.;  
 RT "Cloning and expression of a Streptomyces plicatus chitinase  
 (chitinase-63) in Escherichia coli."  
 RU J. Biol. Chem. 263:443-447(1988).  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-  
 acetyl-D-glucosamine polymers of chitin.  
 CC -1- INDUCTION: By chitin.  
 CC -1- SIMILARITY: Belongs to chitinase class II (family 18 of glycosyl  
 hydrolases). Contains 1 fibronectin type III domain.  
 CC -1- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)  
 domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M82804; AAA26720.1; -.  
 DR PIR; M18397; AAA26717.1; -.  
 DR PIR; JH0573; JH0573.  
 DR HSPD; P07986; 1EXG.  
 DR InterPro; IPR001919; Bag\_cellose\_bind.  
 DR InterPro; IPR008965; Cellul\_bind.  
 DR InterPro; IPR002048; EF-hand.  
 DR InterPro; IPR008957; FN III-like.  
 DR InterPro; IPR003961; FN III.  
 DR InterPro; IPR001223; Glyco\_hydro\_18.  
 DR InterPro; IPR001579; Glyco\_hydro\_18AS.  
 DR Pfam; PF00553; CBM\_2; 1.  
 DR Pfam; PF00041; fn3; 1.  
 DR Pfam; PF00704; Glyco\_hydro\_18; 1.  
 DR ProDom; PD000471; Glyco\_hydro\_18; 1.



```

DR SMART; SM00637; CBD_II; 1.
DR SMART; SM00637; FN3; 1.
DR SMART; SM00637; Glyco_18; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS00561; CBD_BACTERIAL; 1.
DR PROSITE; PS01095; CHITINASE_18; 1.
KW Hydroxylase; Glycosidase; Chitin degradation; Chitin-binding; Signal.
FT SIGNAL 1 30
FT CHAIN 31 610
FT DOMAIN 35 140
FT DOMAIN 148 229
FT DOMAIN 236 610
FT ACT SITE 383 383
SQ SEQUENCE 610 AA; 63974 MW; 6A202EF361CDD500 CRC64;

Query Match 57.8%; Score 48; DB 1; Length 610;
Best Local Similarity 66.7%; Pred. No. 0.98;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 DKLWGVFPYGR 12
DB 498 DKLIGIGFYGR 509

```

## RESULT 15

```

CHIT_STRLI ID CHIT_STRLI STANDARD; PRT; 619 AA.
AC P36909;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Chitinase C precursor (EC 3.2.1.14).
GN CHIC.
OS Streptomyces lividans.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1916;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=93294525; PubMed=8515228;
RX Fujii T., Miyashita K.;
RT "Multiple domain structure in a chitinase gene (chic) of Streptomyces
RT lividans."
RL J. Gen. Microbiol. 139:677-686(1993).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-
CC acetyl-D-glucosamine polymers of chitin.
CC -1- INDUCTION: By chitin.
CC -1- SIMILARITY: Belongs to chitinase class II (family 18 of glycosyl
CC hydrolases).
CC -1- SIMILARITY: Contains 1 fibronectin type III domain.
CC -1- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)
CC domain.
CC
CC This SMARTS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D12647; BAA02168.1; -.
CC HSSP; P07986; IEXG.
CC InterPro; IPR001919; Bac_celose-bind.
CC InterPro; IPR008965; Cellul_bind.
CC InterPro; IPR002048; EF-hand.
CC InterPro; IPR008957; FN_III-like.
CC InterPro; IPR003961; FN_III.
CC InterPro; IPR001223; Glyco_hydro_18.
CC InterPro; IPR001579; Glyco_hydro_18as.
CC Pfam; PF00553; CBM_2; 1.
CC Pfam; PF00041; fn3; 1.

```

```

DR Pfam; PF00704; Glyco_hydro_18; 1.
DR PRODOM; PD000471; Glyco_hydro_18; 1.
DR SMART; SM00637; CBD_II; 1.
DR SMART; SM00637; FN3; 1.
DR SMART; SM00637; Glyco_18; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS00561; CBD_BACTERIAL; 1.
DR PROSITE; PS01095; CHITINASE_18; 1.
KW Hydroxylase; Glycosidase; Chitin degradation; Chitin-binding; Signal.
FT SIGNAL 1 30
FT CHAIN 31 619
FT DOMAIN 35 140
FT DOMAIN 148 230
FT DOMAIN 240 619
FT ACT SITE 382 382
SQ SEQUENCE 619 AA; 65200 MW; A23CE5B3C5D6F21 CRC64;

Query Match 57.8%; Score 48; DB 1; Length 619;
Best Local Similarity 66.7%; Pred. No. 0.99;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 DKLWGVFPYGR 12
DB 497 DKLIGIGFYGR 508

```

Search completed: March 22, 2004, 06:53:02  
Job time : 1.65203 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:39:53 ; Search time 3.42079 Seconds

(without alignments)  
1568.003 Million cell updates/sec

Title: US-09-662-293-10

Sequence: 1 DRLVNGVFPYGRASIE 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :  
1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP rodent:\*  
12: SP virus:\*  
13: SP vertebrate:\*  
14: SP unclassified:\*  
15: SP xvirus:\*  
16: SP bacteriap:\*  
17: SP archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	80	96.4	555	5	Q9U6R7 dermatophag
2	65	78.3	592	2	Q84F90 bacillus j1
3	62	74.7	499	2	Q8KWS2 serratia 11
4	62	74.7	674	2	Q9FAC8 bacillus ce
5	62	74.7	674	2	Q8RSX8 bacillus th
6	62	74.7	674	16	Q81287 bacillus a
7	62	74.7	674	16	Q81IF9 bacillus c
8	62	74.7	676	2	Q93AD3 bacillus th
9	62	74.7	688	2	Q8KUT8 bacillus ce
10	62	74.7	688	2	Q8KUT3 bacillus th
11	62	74.7	688	2	Q8KQ05 bacillus th
12	62	74.7	688	2	Q84FN2 bacillus th
13	61	73.5	499	2	Q54276 serratia ma
14	61	73.5	522	2	Q59929 serratia ma
15	60	72.3	268	5	Q8MS14 drosophila
16	60	72.3	460	5	Q9W2M7 drosophila

17	60	72.3	496	2	Q9RG51	Q9RG51 bacillus ci
18	60	72.3	554	16	Q8A793	Q8A793 bacteroides
19	60	72.3	596	2	Q85500	Q85500 bacillus su
20	59	71.1	572	5	Q26042	Q26042 penaeus jap
21	59	71.1	620	5	Q9Y0D4	Q9Y0D4 penaeus mon
22	58	69.9	484	5	Q9W092	Q9W092 drosophila
23	58	69.9	543	5	Q8GR93	Q8GR93 bombyx mori
24	58	69.9	543	5	Q8GV05	Q8GV05 bombyx mori
25	58	69.9	544	5	Q9GQC4	Q9GQC4 bombyx mori
26	58	69.9	552	5	Q9GQC4	Q9GQC4 spodoptera
27	58	69.9	553	5	P17311	P17311 hyalantaria
28	58	69.9	557	5	Q8MTX0	Q8MTX0 choristoneu
29	58	69.9	565	5	P90710	P90710 bombyx mori
30	58	69.9	565	5	Q9GPG9	Q9GPG9 bombyx mand
31	58	69.9	566	5	Q8WR52	Q8WR52 bombyx mori
32	58	69.9	574	5	Q17411	Q17411 aedes aegypt
33	58	69.9	595	5	Q9VPR3	Q9VPR3 drosophila
34	58	69.9	665	2	Q48373	Q48373 janthinobac
35	57	68.7	534	2	Q8VUC6	Q8VUC6 pseudomonas
36	56	67.5	633	5	Q22468	Q22468 caenorhabdi
37	56	67.5	699	2	Q48494	Q48494 kurtina zop
38	56	67.5	717	2	Q9KHB3	Q9KHB3 bacillus ci
39	56	67.5	4498	5	Q9W223	Q9W223 drosophila
40	55	66.3	572	17	Q9HR83	Q9HR83 halobacteri
41	55	66.3	599	16	Q9KED7	Q9KED7 bombyx mori
42	55	66.3	1080	5	Q9BLI6	Q9BLI6 bombyx mori
43	54	65.1	688	5	Q8MS85	Q8MS85 drosophila
44	53	63.9	143	2	Q8VQH0	Q8VQH0 gamma-prote
45	53	63.9	145	2	Q8RNA1	Q8RNA1 gamma-prote

## ALIGNMENTS

RESULT 1  
Q9U6R7 PRELIMINARY; PRT; 555 AA.  
ID Q9U6R7  
AC Q9U6R7  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DR 98KDa HDM allergen.  
OS Dermatophagoides farinae (House-dust mite).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
OC Acariformes; Sarcoptiformes; Astigmata; Psoroptida; Analgoidea;  
OC Pyroglyphidae; Dermatophagoides.  
OX NCBI\_TaxId=6954;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Weber E.R., Hunter S., Steedman K., McCall C.;  
RT "Cloning and Characterization of a 98 KDa Allergen from  
RT Dermatophagoides farinae."  
RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF178772; AAC52672.1;  
DR GO; GO:0005576; C:extracellular; IEA.  
DR GO; GO:0008061; F:chitin binding; IEA.  
DR GO; GO:0016798; F:hydrolyase activity; acting on glycosyl bonds; IEA.  
DR GO; GO:0009575; P:carbohydrate metabolism; IEA.  
DR GO; GO:0006030; P:chitin metabolism; IEA.  
DR InterPro; IPR002557; Chitin\_bind\_Pera.  
DR InterPro; IPR001223; Glyco\_hydro\_18.  
DR InterPro; IPR001579; Glyco\_hydro\_18a.  
DR Pfam; PF00704; Glyco\_hydro\_18; 1.  
DR Prodom; PD000471; Glyco\_hydro\_18; 1.  
DR SMART; SM00494; ChitBD2\_1; 1.  
DR SMART; SM00636; Glyco\_18; 1.  
DR PROSITE; PS01095; CHITINASE\_18; 1.  
KW Glycosidase; Hydrolase.  
SQ SEQUENCE 555 AA; 63238 MW; 0E4564A1A459B30B CRC64;  
Query Match 96.4%; Score 80; DB 5; Length 555;  
Best Local Similarity 94.1%; Pred. No. 1.1e-05;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Cy 1 DKLWGVFPYGRAXSIE 17
Db 273 DKLWGVFPYGRAXSIE 289

RESULT 2
084F90 PRELIMINARY; PRT; 592 AA.
ID 084F90;
AC 084F90;
DT 01-JUN-2003 (TREMblrel. 24, Created)
DT 01-JUN-2003 (TREMblrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Chitinase.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_Taxid=1402;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PR-1;
RA Rutdeekulthamrong P., Pichyangkura R.;
RT "Chitinase and chitodextrinase genes from Bacillus licheniformis PR-1."
RU Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY205293; MA02144.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0030246; F:carbohydrate binding; IEA.
DR GO; GO:0004553; F:hydrolyase activity; IEA.
DR GO; GO:0016787; F:hydrolyase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR003610; CBM_5_12.
DR InterPro; IPR003962; FnIII_subd.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR001223; Glyco_hydro_18.
DR InterPro; IPR001579; Glyco_hydro_18AS.
DR Pfam; PF02839; CBM_5_12; 1.
DR Pfam; PF00041; fn3; 1.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR PRINTS; PR00014; FNTYPEIII.
DR ProDom; PD000471; Glyco_hydro_18; 2.
DR SMART; SM00495; ChtBD3; 1.
DR SMART; SM0060; FN3; 1.
DR SMART; SM00636; Glyco_18; 1.
DR PROSITE; PS01095; CHITINASE_18; 1.
SQ SEQUENCE 592 AA; 6516 MW; F936B24212D220 CRC64;

Query Match 78.3%; Score 65; DB 2; Length 592;
Best Local Similarity 68.8%; Pred. No. 0.0059;
Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```

```

RT "Analysis of the chb gene of Serratia liquefaciens."
RL J. Biotechnol. 80:277-283(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Moykovich A.E., Selvaraj G., Khachatourians G.G.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF399871; AA03597.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0030246; F:carbohydrate binding; IEA.
DR GO; GO:0004553; F:hydrolyase activity; IEA.
DR GO; GO:0016787; F:hydrolyase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR003610; CBM_5_12.
DR InterPro; IPR001223; Glyco_hydro_18.
DR InterPro; IPR001579; Glyco_hydro_18AS.
DR Pfam; PF02839; CBM_5_12; 1.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR ProDom; PD000471; Glyco_hydro_18; 1.
DR SMART; SM00495; ChtBD3; 1.
DR SMART; SM00636; Glyco_18; 1.
DR PROSITE; PS01095; CHITINASE_18; 1.
SQ SEQUENCE 499 AA; 55441 MW; 5FC31028003C9419 CRC64;

Query Match 74.7%; Score 62; DB 2; Length 499;
Best Local Similarity 84.6%; Pred. No. 0.017;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

Cy 1 DKLWGVFPYGRAXSIE 13
Db 283 NKIWMGVFPYGRAXSIE 295

RESULT 4
09FAC8 PRELIMINARY; PRT; 674 AA.
ID 09FAC8;
AC 09FAC8;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Chitinase B.
OS Bacillus cereus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_Taxid=1396;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CH;
RA Mabuchi N., Araki Y.;
RL "Chitinase B."
RT Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
DR EMBL; AB041932; BAB16891.1; -.
DR HSP; P07254; ICTN.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004553; F:hydrolyase activity; IEA.
DR GO; GO:0004812; F:RNA ligase activity; IEA.
DR GO; GO:0006418; P:amino acid activation; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; Bac_celose-bind.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR001223; Glyco_hydro_18.
DR InterPro; IPR001579; Glyco_hydro_18AS.
DR InterPro; IPR001412; CRNA-syn_1.
DR Pfam; PF00553; CBM_2; 1.
DR Pfam; PF00041; fn3; 1.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR ProDom; PD000471; Glyco_hydro_18; 2.
DR SMART; SM00637; CBD_II; 1.
DR SMART; SM0060; FN3; 1.
DR SMART; SM00636; Glyco_18; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.

```

```

DR PROSITE; PS01095; CHITINASE_18; 1.
KM Glycosidase; Hydrolase.
SQ SEQUENCE 674 AA; 74262 MW; 66BE9FE80D660561 CRC64;

Query Match 74.7%; Score 62; DB 2; Length 674;
Best Local Similarity 91.7%; Pred. No. 0.023;
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DKLVGVPPFYGR 12
DQ 330 DKLVGVPPFYGR 341

RESULT 5
Q8RSX8 PRELIMINARY; PRT; 674 AA.
AC Q8RSX8
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Chitinase.
OS Bacillus thuringiensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1428;
RN [1]
RP SEQUENCE FROM N.A.
RA Sha L., Guan X., Li L.;
RT "The gene cloning of Bacillus thuringiensis chitinase.";
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY074882; AAL71886.2; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004553; F:hydrolase activity; hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004812; F:crRNA ligase activity; IEA.
DR GO; GO:0006418; F:amino acid activation; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; Bac celose-bind.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR003961; FN_III-like.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR001223; Glyco_hydro_18.
DR InterPro; IPR001579; Glyco_hydro_18AS.
DR InterPro; IPR001412; rRNA-synt_1.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR ProDom; PD000471; Glyco_hydro_18; 2.
DR SMART; SM00637; CBD_II; 1.
DR SMART; SM00060; FN3; 1.
DR SMART; SM00636; Glyco_18; 1.
DR SMART; SM00636; Glyco_18; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
DR PROSITE; PS01095; CHITINASE_18; 1.
SQ SEQUENCE 674 AA; 74209 MW; 8BE1C978751D17B CRC64;

Query Match 74.7%; Score 62; DB 2; Length 674;
Best Local Similarity 91.7%; Pred. No. 0.023;
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DKLVGVPPFYGR 12
DQ 330 DKLVGVPPFYGR 341

RESULT 6
Q81287 PRELIMINARY; PRT; 674 AA.
AC Q81287
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Chitinase B.
GN BA0385.
OS Bacillus anthracis (strain Ames).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=198094;

RN [1]
RP SEQUENCE FROM N.A.
RA Read T.D., Peterson S.N., Tourasse N., Baillye L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtapple E.K., Oksrad O.A., Helgason E., Ralston J., Wu M.,
RA Kolman J.F., Bearan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Hart D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouli H.M., Radune D.,
RA Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plant R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.B.,
RA Fraser C.M.;
RT "The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria.";
RL Nature 423:81-86(2003).
DR EMBL; AE017025; AAP2415.1; -.
DR TRIM; BA0385; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004553; F:hydrolase activity; hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004812; F:crRNA ligase activity; IEA.
DR GO; GO:0006418; F:amino acid activation; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; Bac celose-bind.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR003961; FN_III-like.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR001223; Glyco_hydro_18.
DR InterPro; IPR001579; Glyco_hydro_18AS.
DR InterPro; IPR001412; rRNA-synt_1.
DR Pfam; PF00553; CBM_2; 1.
DR Pfam; PF00041; FN3; 1.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR ProDom; PD000471; Glyco_hydro_18; 2.
DR SMART; SM00637; CBD_II; 1.
DR SMART; SM00060; FN3; 1.
DR SMART; SM00636; Glyco_18; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
DR PROSITE; PS01095; CHITINASE_18; 1.
KM Complete proteome.
SQ SEQUENCE 674 AA; 74161 MW; 211E3206995524A6 CRC64;

Query Match 74.7%; Score 62; DB 16; Length 674;
Best Local Similarity 91.7%; Pred. No. 0.023;
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DKLVGVPPFYGR 12
DQ 330 DKLVGVPPFYGR 341

RESULT 7
Q811F9 PRELIMINARY; PRT; 674 AA.
AC Q811F9
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Endochitinase (EC 3.2.1.14).
GN BC0429.
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=226900;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=22608415; PubMed=12721630;
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelson B.,
RA Kapural V., Bhattacharya A., Reznik G., Mikhailova N., Lapidus A.,
RA Chu L., Mazur M., Goldsman E., Larsen N., D'Souza M., Watkins T.,
RA Greckin Y., Pasch G., Haselkorn R., Fomstein M., Ehrlich S.D.,
RA Overbeek R., Kyrides N.;

```

RT "Genome sequence of *Bacillus cereus* and comparative analysis with  
 RT *Bacillus anthracis*." ;  
 RL Nature 423:87-91(2003).  
 DR EMBL; AE016999; AAP07469.1; -.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008843; F:endochitinase activity; IEA.  
 DR GO; GO:0004812; F:RNA ligase activity; IEA.  
 DR GO; GO:0006418; P:amino acid activation; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR01919; Bac celose-bind.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR008957; FN\_III-like.  
 DR InterPro; IPR01223; Glyco\_hydro\_18.  
 DR InterPro; IPR01579; Glyco\_hydro\_18A.  
 DR InterPro; IPR01412; tRNA-synt\_1.  
 DR Pfam; PF00553; CBM\_2; 1.  
 DR Pfam; PF00041; fn3; 1.  
 DR Pfam; PF00704; Glyco\_hydro\_18; 1.  
 DR ProDom; PD000471; Glyco\_hydro\_18; 2.  
 DR SMART; SMO0637; CBD\_II; 1.  
 DR SMART; SMO060; FN3; 1.  
 DR SMART; SMO0636; Glyco\_18; 1.  
 DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_I; 1.  
 DR PROSITE; PS01095; CHITINASE\_18; 1.  
 DR GlycoSide; HydroLase; Complete proteome.  
 SQ SEQUENCE 674 AA; 74237 MW; 252EF2757FB3FB18 CRC64;  
 Query Match 74.7%; Score 62; DB 16; Length 674;  
 Best Local Similarity 91.7%; Pred. No. 0.023;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DKLVMGVPPFYGR 12  
 Db 330 DKLVLGVPPYGR 341  
 RESULT 8  
 Q93AD3 PRELIMINARY; PRT; 676 AA.  
 AC Q93AD3;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Chitinase (EC 3.2.1.14).  
 GN CHI.  
 OS *Bacillus thuringiensis*.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxId=1428;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Baroja-Corona E., Velazquez-Robledo R., Nieto-Mazocco E.;  
 RT "Cloning of a chitinase gene from a Mexican strain of *Bacillus*  
 RT *thuringiensis*." ;  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.  
 DR EMBL; AF424979; BAI17867.1; -.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008843; F:endochitinase activity; IEA.  
 DR GO; GO:0004812; F:RNA ligase activity; IEA.  
 DR GO; GO:0006418; P:amino acid activation; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR01919; Bac celose-bind.  
 DR InterPro; IPR008957; FN\_III-like.  
 DR InterPro; IPR01223; Glyco\_hydro\_18.  
 DR InterPro; IPR01579; Glyco\_hydro\_18A.  
 DR InterPro; IPR01412; tRNA-synt\_1.  
 DR Pfam; PF00553; CBM\_2; 1.  
 DR Pfam; PF00041; fn3; 1.  
 DR ProDom; PD000471; Glyco\_hydro\_18; 1.  
 DR ProDom; PD000471; Glyco\_hydro\_18; 2.

DR SMART; SMO0637; CBD\_II; 1.  
 DR SMART; SMO060; FN3; 1.  
 DR SMART; SMO0636; Glyco\_18; 1.  
 DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_I; 1.  
 DR PROSITE; PS01095; CHITINASE\_18; 1.  
 DR GlycoSide; HydroLase.  
 SQ SEQUENCE 676 AA; 74469 MW; A7B7DAPB621516C9 CRC64;  
 Query Match 74.7%; Score 62; DB 2; Length 676;  
 Best Local Similarity 91.7%; Pred. No. 0.023;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DKLVMGVPPFYGR 12  
 Db 332 DKLVLGVPPYGR 343  
 RESULT 9  
 Q8KVU8 PRELIMINARY; PRT; 688 AA.  
 AC Q8KVU8;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Chitinase CW (EC 3.2.1.14).  
 GN CHICW.  
 OS *Bacillus cereus*.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxId=1396;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Chen C.-Y., Huang C.-U., Wang T.-K.;  
 RT "Cloning of a chitinase gene from an antagonistic strain of *Bacillus*  
 RT *cereus*." ;  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.  
 DR EMBL; AF416570; AAM48520.1; -.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008843; F:endochitinase activity; IEA.  
 DR GO; GO:0004812; F:RNA ligase activity; IEA.  
 DR GO; GO:0006418; P:amino acid activation; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR01919; Bac celose-bind.  
 DR InterPro; IPR008957; FN\_III-like.  
 DR InterPro; IPR01223; Glyco\_hydro\_18.  
 DR InterPro; IPR01579; Glyco\_hydro\_18A.  
 DR InterPro; IPR01412; tRNA-synt\_1.  
 DR Pfam; PF00553; CBM\_2; 1.  
 DR Pfam; PF00041; fn3; 1.  
 DR Pfam; PF00704; Glyco\_hydro\_18; 1.  
 DR ProDom; PD000471; Glyco\_hydro\_18; 2.  
 DR SMART; SMO0637; CBD\_II; 1.  
 DR SMART; SMO060; FN3; 1.  
 DR SMART; SMO0636; Glyco\_18; 1.  
 DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_I; 1.  
 DR PROSITE; PS01095; CHITINASE\_18; 1.  
 DR HydroLase; GlycoSide.  
 SQ SEQUENCE 688 AA; 75859 MW; 69AB97F728A292E6 CRC64;  
 Query Match 74.7%; Score 62; DB 2; Length 688;  
 Best Local Similarity 91.7%; Pred. No. 0.024;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DKLVMGVPPFYGR 12  
 Db 344 DKLVLGVPPYGR 355  
 RESULT 10  
 Q8KMY3

```

ID 08KXV3 PRELIMINARY; PRT; 688 AA.
AC 08KXV3;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Chitinase.
GN CHI.
OS Bacillus thuringiensis (subsp. sotto).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxId=29340;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhong W.F., Yan W.Z., Jiang L.H., Cai P.Z.;
RT "Cloning and sequence analysis of chitinase gene from Bacillus
   thuringiensis."
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTRAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
DR GO: GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO: GO:0004553; F:ATP binding; IEA.
DR GO: GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO: GO:0016787; F:hydrolase activity; IEA.
DR GO: GO:0004812; F:RNA ligase activity; IEA.
DR GO: GO:0004418; F:amino acid activation; IEA.
DR GO: GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro: IPR001919; Bac celose-bind.
DR InterPro: IPR003961; FN_III-like.
DR InterPro: IPR008957; FN_III-like.
DR InterPro: IPR001223; Glyco_hydro_18.
DR InterPro: IPR001579; Glyco_hydro_18AS.
DR Pfam: PF00553; CBM_2; 1.
DR Pfam: PF00704; Glyco_hydro_18; 1.
DR Pfam: PF00704; Glyco_hydro_18; 2.
DR SMART: SM00637; CBD_II; 1.
DR SMART: SM00637; FN3; 1.
DR SMART: SM00636; Glyco_18; 1.
DR PROSITE: PS00178; AA_TRNA_LIGASE_I; 1.
DR PROSITE: PS01095; CHITINASE_18; 1.
DR Hydrolase.
SQ SEQUENCE 688 AA; 75751 MW; CSEF68EA43897A9E CRC64;

Query Match 74.7%; Score 62; DB 2; Length 688;
Best Local Similarity 91.7%; Pred. No. 0.024;
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DKLVGVPEYGR 12
Db 344 DKLVGVPEYGR 355

```

```

DR GO: GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO: GO:0016787; F:hydrolase activity; IEA.
DR GO: GO:0004812; F:RNA ligase activity; IEA.
DR GO: GO:0004418; F:amino acid activation; IEA.
DR GO: GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro: IPR001919; Bac celose-bind.
DR InterPro: IPR008957; Cellul_bind.
DR InterPro: IPR003961; FN_III-like.
DR InterPro: IPR008957; FN_III-like.
DR InterPro: IPR001223; Glyco_hydro_18.
DR InterPro: IPR001579; Glyco_hydro_18AS.
DR Pfam: PF00553; CBM_2; 1.
DR Pfam: PF00704; Glyco_hydro_18; 1.
DR Pfam: PF00704; Glyco_hydro_18; 2.
DR SMART: SM00637; CBD_II; 1.
DR SMART: SM00637; FN3; 1.
DR SMART: SM00636; Glyco_18; 1.
DR PROSITE: PS00178; AA_TRNA_LIGASE_I; 1.
DR PROSITE: PS01095; CHITINASE_18; 1.
DR Hydrolase.
SQ SEQUENCE 688 AA; 75788 MW; 0FCDDC71760D1C7 CRC64;

Query Match 74.7%; Score 62; DB 2; Length 688;
Best Local Similarity 91.7%; Pred. No. 0.024;
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DKLVGVPEYGR 12
Db 344 DKLVGVPEYGR 355

RESULT 12
ID 084FN2 PRELIMINARY; PRT; 688 AA.
AC 084FN2;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Chitinase.
GN CHI.
OS Bacillus thuringiensis (subsp. kurstaki).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxId=29339;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhong W.F., Jiang L.H., Yan W.Z., Cai P.Z., Xiang Y.W., Zhang Z.X.;
RT "Characterization of chitinase gene cloned from Bacillus thuringiensis
   kurstaki."
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR GO: GO:0004553; F:ATP binding; IEA.
DR GO: GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO: GO:0016787; F:hydrolase activity; IEA.
DR GO: GO:0004812; F:RNA ligase activity; IEA.
DR GO: GO:0004418; F:amino acid activation; IEA.
DR GO: GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro: IPR001919; Bac celose-bind.
DR InterPro: IPR008957; Cellul_bind.
DR InterPro: IPR003961; FN_III-like.
DR InterPro: IPR008957; FN_III-like.
DR InterPro: IPR001223; Glyco_hydro_18.
DR InterPro: IPR001579; Glyco_hydro_18AS.
DR Pfam: PF00553; CBM_2; 1.
DR Pfam: PF00704; Glyco_hydro_18; 1.
DR Pfam: PF00704; Glyco_hydro_18; 2.
DR SMART: SM00637; CBD_II; 1.
DR SMART: SM00637; FN3; 1.
DR SMART: SM00636; Glyco_18; 1.
DR PROSITE: PS00178; AA_TRNA_LIGASE_I; 1.
DR PROSITE: PS01095; CHITINASE_18; 1.

```

DR PROSITE: PS01095; CHITINASE\_18; 1.  
 SQ SEQUENCE 688 AA; 75820 MW; F9D552062575AAB4 CRC64;  
 Query Match 74.7%; Score 62; DB 2; Length 688;  
 Best Local Similarity 91.7%; Pred. No. 0.024;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DKLVGVPPFYGR 12  
 Db 344 DKLVGVPPFYGR 355

RESULT 13  
 ID Q54276 PRELIMINARY; PRT; 499 AA.  
 AC Q54276;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Chitinase.  
 GN CHIB.  
 OS Serratia marcescens.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Serratia.  
 OX NCBI\_TaxID=615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95202070; PubMed=7894703;  
 RA Bruberg M.B., Elsjink V.G.H., Haandrikman A.J., Venema G., Nes I.F.;  
 RT "Chitinase B from *Serratia marcescens* BUL200 is exported to the  
 RT periplasm without processing.";  
 RL Microbiology 141:123-131(1995).  
 DR EMBL: Z36295; CA85292.1; -  
 DR PIR: S52422; S52422.  
 DR PDB: 1E15; 18-AUG-00.  
 DR PDB: 1E6N; 22-JUN-01.  
 DR PDB: 1E6P; 22-JUN-01.  
 DR PDB: 1E6R; 22-JUN-01.  
 DR PDB: 1E6Z; 22-JUN-01.  
 DR PDB: 1G0I; 15-NOV-01.  
 DR PDB: 1GPF; 31-OCT-02.  
 DR GO: GO:0005576; C:extracellular; IEA.  
 DR GO: GO:0030246; F:carbohydrate binding; IEA.  
 DR GO: GO:0004553; F:hydrolyase activity; hydrolyzing O-glycosyl . . .; IEA.  
 DR GO: GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro: IPR003610; CBM\_5\_12.  
 DR InterPro: IPR001223; Glyco\_hydro\_18.  
 DR InterPro: IPR001579; Glyco\_hydro\_18AS.  
 DR Pfam: PF02839; CBM\_5\_12; 1.  
 DR Pfam: PF00704; Glyco\_hydro\_18; 1.  
 DR Prodom: PD000471; Glyco\_hydro\_18; 1.  
 DR SMART: SM00495; ChnBD3; 1.  
 DR SMART: SM00636; Glyco\_18; 1.  
 DR PROSITE: PS01095; CHITINASE\_18; 1.  
 KM Glycosidase; Hydrolase.  
 SQ SEQUENCE 499 AA; 55469 MW; 58C933A9064D526B CRC64;

Query Match 73.5%; Score 61; DB 2; Length 499;  
 Best Local Similarity 91.7%; Pred. No. 0.026;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KLVGVPPFYGR 13  
 Db 284 KLVGVPPFYGR 295

RESULT 14  
 ID Q59929 PRELIMINARY; PRT; 522 AA.  
 AC Q59929;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Chitinase precursor (EC 3.2.1.14).  
 OS Serratia marcescens.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Serratia.  
 OX NCBI\_TaxID=615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97372534; PubMed=9228754;  
 RA Gal S.W., Choi J.Y., Kim C.Y., Cheong Y.H., Choi Y.J., Bahk J.D.,  
 RA Lee S.Y., Cho M.J.;  
 RT "Isolation and characterization of the 54-kDa and 22-kDa chitinase  
 RT genes of *Serratia marcescens* KCTC2172.";  
 RL FEMS Microbiol. Lett. 151:197-204(1997).  
 DR EMBL: U38484; AAC37122.1; -  
 DR HSBP; P07254; 1CTN.  
 DR GO: GO:0005576; C:extracellular; IEA.  
 DR GO: GO:0030246; F:carbohydrate binding; IEA.  
 DR GO: GO:0008843; F:endochitinase activity; IEA.  
 DR GO: GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro: IPR003610; CBM\_5\_12.  
 DR InterPro: IPR001223; Glyco\_hydro\_18.  
 DR Pfam: PF02839; CBM\_5\_12; 1.  
 DR Pfam: PF00704; Glyco\_hydro\_18; 1.  
 DR Prodom: PD000471; Glyco\_hydro\_18; 1.  
 DR SMART: SM00495; ChnBD3; 1.  
 DR SMART: SM00636; Glyco\_18; 1.  
 KW Glycosidase; Hydrolase; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 1 522  
 SQ SEQUENCE 522 AA; 58107 MW; D310295B296AFT0 CRC64;

Query Match 73.5%; Score 61; DB 2; Length 522;  
 Best Local Similarity 91.7%; Pred. No. 0.027;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KLVGVPPFYGR 13  
 Db 307 KLVGVPPFYGR 318

RESULT 15  
 ID Q8MS14 PRELIMINARY; PRT; 268 AA.  
 AC Q8MS14;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE GH1872P.  
 GN CG9357.  
 OS Drosophila melanogaster (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Diptera;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RA Stapleton M., Brokstein P., Hong L., Aghayani A., Carlson J.,  
 RA Champagne M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
 RA George R., Gonzalez C.M., Guarin H., Kronmiller B., Li P., Liao G.,  
 RA Mirza A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,  
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
 RA Ceoliker S.;  
 RT Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY118784; AAM50644.1; -  
 DR FlyBase: FBgn034580; CG9357.  
 DR GO: GO:0005576; C:extracellular; IEA.  
 DR GO: GO:0008061; F:chitin binding; IEA.  
 DR GO: GO:0016787; F:hydrolyase activity; IEA.  
 DR GO: GO:0006030; P:chitin metabolism; IEA.  
 DR GO: GO:0008152; P:metabolism; IEA.  
 DR GO: GO:0006457; P:protein folding; IEA.  
 DR InterPro: IPR002557; Chitin\_bind\_Petr.

DR InterPro: IPR001179; FKBP\_PPIase.  
 DR InterPro: IPR001223; Glyco\_hydro\_18.  
 DR Pfam: PF01607; CBM\_14; 1.  
 DR Pfam: PF00704; Glyco\_hydro\_18; 1.  
 DR ProDom: PD000471; Glyco\_hydro\_18; 1.  
 DR SMART: SM00494; ChcBD2; 1.  
 DR SMART: SM00636; Glyco\_18; 1.  
 DR PROSITE: PS00453; FKBP\_PPIASE\_1; 1.  
 SQ SEQUENCE 268 AA; 29678 MW; 5DEBDC9903CB2DIE CRC64;

Query Match 72.3%; Score 60; DB 5; Length 268;  
 Best Local Similarity 56.2%; Pred. No. 0.02;  
 Matches 9; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKLVGVFFYGRAXSI 16  
 :||:|||||: :  
 Db 63 EKLIVGVFFYGRSFTL 78

Search completed: March 22, 2004, 06:59:21  
 Job time : 4.42079 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: March 22, 2004, 06:30:23 ; Search time 5.16898 Seconds  
(without alignments)  
929.256 Million cell updates/sec

Title: US-09-662-293-10

Sequence: 1 DKLMGVFPYGRAXSIE 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	81	97.6	17	AAV52519	AAV52519 House dus
2	81	97.6	17	AAU96323	AAU96323 Der HMW-m
3	80	96.4	490	AAV52535	AAV52535 D. pteron
4	80	96.4	490	AAU96339	AAU96339 Der HMW-m
5	80	96.4	509	AAV52533	AAV52533 D. pteron
6	80	96.4	509	AAU96337	AAU96337 Der HMW-m
7	80	96.4	509	AAU96338	AAU96338 Der HMW-m
8	80	96.4	536	AAV52525	AAV52525 House dus
9	80	96.4	536	AAU96339	AAU96339 Der HMW-m
10	80	96.4	555	AAV52523	AAV52523 House dus
11	80	96.4	555	AAU96327	AAU96327 Der HMW-m
12	80	96.4	555	AAU96328	AAU96328 Der HMW-m
13	80	72.3	460	ABE64366	ABE64366 Drosophila
14	59	71.1	489	AAE28202	AAE28202 Flea chit
15	59	71.1	559	AAE28199	AAE28199 Flea PCFC
16	59	71.1	583	AAE28197	AAE28197 Flea chit
17	58	69.9	484	ABE58787	ABE58787 Drosophila
18	58	69.9	553	ABP72626	ABP72626 Hyphantri
19	58	69.9	554	AAW01824	AAW01824 Manduca s
20	58	69.9	554	AAW01824	AAW01824 Manduca s
21	58	69.9	554	ABP72619	ABP72619 Manduca s
22	58	69.9	565	ABP72625	ABP72625 Bombyx mo
23	58	69.9	574	ABP72635	ABP72635 Aedes aeg
24	58	69.9	595	ABR11737	ABR11737 Drosophila
25	56	67.5	458	ABE66690	ABE66690 Drosophila

26	56	67.5	635	5	AAE28203	AAE28203 Flea chit
27	56	67.5	699	6	ABP72622	ABP72622 Bacillus
28	56	67.5	4498	6	ABE58595	ABE58595 Drosophila
29	54	65.1	305	4	ABE66689	ABE66689 Drosophila
30	54	65.1	508	4	ABE66399	ABE66399 Drosophila
31	52	62.7	371	2	AAE70025	AAE70025 Tobacco c
32	52	62.7	377	2	AAE70029	AAE70029 Tobacco c
33	51	61.4	718	2	AAE73992	AAE73992 Murine cv
34	50	60.2	383	2	AAE26751	AAE26751 Bovine wh
35	50	60.2	383	2	AAE03442	AAE03442 G338k pro
36	49	59.0	227	7	ADCI4213	ADCI4213 Human enz
37	49	59.0	305	7	ADCI4212	ADCI4212 Human enz
38	49	59.0	383	5	ABE66297	ABE66297 Human ova
39	49	59.0	383	6	ABE6651	ABE6651 Lung carc
40	49	59.0	383	6	ABE69725	ABE69725 Protein d
41	49	59.0	385	7	ADE58277	ADE58277 Human Pro
42	49	59.0	399	4	AAU00962	AAU00962 F. venena
43	49	59.0	424	2	AAE63441	AAE63441 Trichoder
44	49	59.0	424	3	AAE68728	AAE68728 Amino aci
45	49	59.0	424	3	AAE51402	AAE51402 T. harzia

## ALIGNMENTS

RESULT 1	AAV52519	standard; peptide; 17 AA.
XX	AAV52519;	
AC	AAV52519;	
XX	22-FEB-2000	(first entry)
DT	22-FEB-2000	(first entry)
XX	House dust mite allergen protein (map) A/B fragment map (9).	
DE	House dust mite allergen protein (map) A/B fragment map (9).	
XX	Mite allergen protein; map; high molecular weight; HMW-map; allergy;	
KW	house dust mite; IGB; immunoglobulin E; allergen; mapA; mapB;	
KW	hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;	
KW	canine; veterinary; antibody; vaccine; immunisation.	
XX	Dermatophagoides farinae.	
OS	Dermatophagoides farinae.	
XX	Key	Location/Qualifiers
FT	Misc-difference 14	/label= Xaa
FT	Misc-difference 14	/note= "Xaa = any amino acid"
FT	Misc-difference 14	/note= "Xaa = any amino acid"
XX	W09954349-A2.	
XX	28-OCT-1999.	
PD	28-OCT-1999.	
XX	16-APR-1999;	99MO-US008524.
PF	16-APR-1999;	99MO-US008524.
XX	17-APR-1998;	98US-00062013.
PR	13-MAY-1998;	98US-0085295P.
PR	02-SEP-1998;	98US-0098909P.
XX	(HESK-) HESKA CORP.	
PA	(HESK-) HESKA CORP.	
XX	Mccall CA, Hunter SW, Weber ER;	
PI	Mccall CA, Hunter SW, Weber ER;	
XX	WPI, 2000-052700/04.	
DR	WPI, 2000-052700/04.	
XX	Novel high molecular weight Dermatophagoides nucleic acid polypeptides	
PT	used to modify an animals' hypersensitivity to mite allergens.	
FT	used to modify an animals' hypersensitivity to mite allergens.	
PS	Claim 3; Page 70; 154p; English.	
XX	Sequences AAV52510-V52522 represent proteolytic fragments of	
CC	Dermatophagoides farinae high molecular weight mite allergen protein (HMW	
CC	-map) composition. The HMW-map composition was isolated from a D. farinae	
CC	homogenate by gel filtration, with each fraction being analysed for the	
CC	presence of proteins that bound to IGE present in mite-allergic dog	

CC antisera. The HMM-map composition comprises mapA (a 109 kD protein) and  
 CC mapB (98 kD). Mite allergenic proteins and peptides, and nucleic acids  
 CC encoding them, may be used in therapeutic compositions to modify an  
 CC animal's hypersensitivity reaction to mite allergens. Animals that may be  
 CC treated include mammals and birds, especially felines, canines, equines,  
 CC humans, other pets, and work or domestic animals. The proteins or  
 CC fragments may also be used to diagnose allergies via a skin test. The  
 CC proteins and peptides can also be used to raise antibodies, which have a  
 CC variety of potential uses. For example, they can be used as vaccines to  
 CC passively immunise animals against dust mite hypersensitivity, as  
 CC positive controls in test kits and as tools to recover desired dust mite  
 CC allergens from a mixture of proteins

XX SQ Sequence 17 AA:

Query Match 97.6%; Score 81; DB 3; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 2,8e-08;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DKLVMGVPFYGRAXSIE 17  
 |||||  
 1 DKLVMGVPFYGRAXSIE 17

RESULT 2

AAU96323  
 ID AAU96323 standard; peptide; 17 AA.

XX AAU96323;

XX 15-JUL-2002 (first entry)

XX Der HMM-map polypeptide #10.

XX Der HMM-map: American house dust mite; antiallergic; mite; IgE;  
 XX mite allergenic protein; immunoglobulin E; hypersensitivity;  
 XX immunocomplex formation.

XX Dermatophagoides farinae.

XX WO200222807-A2.

XX 21-MAR-2002.

XX 14-SEP-2001; 2001WO-US028730.

XX 14-SEP-2000; 2000US-00662293.

XX (HESK-) HESKA CORP.

XX McCall CA, Hunter SM, Weber ER;

XX WPI; 2002-351888/38.

XX New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HMM-map protein, useful as a vaccine for treating mite allergy.

XX Claim 12; Page 71; 161pp; English.

XX The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HMM-map protein, and its related nucleic  
 CC acid. The Der HMM-map protein is useful for eliciting an immune response  
 CC against Der HMM-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitising a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMM-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig)E or Der HMM-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HMM-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342

CC represent Der HMM-map polypeptides of the invention  
 XX SQ Sequence 17 AA:

Query Match 97.6%; Score 81; DB 5; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 2,8e-08;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DKLVMGVPFYGRAXSIE 17  
 |||||  
 1 DKLVMGVPFYGRAXSIE 17

RESULT 3

AAU52535  
 ID AAU52535 standard; protein; 490 AA.

XX AAU52535;

XX 06-AUG-2003 (revised)  
 DT 22-FEB-2000 (first entry)

XX D. pteronyssinus 98 kD mite allergen protein (map) pDerp98-490.

XX Mite allergen protein; map; high molecular weight; HMM-map; allergy;  
 XX house dust mite; IgE; immunoglobulin E; allergen; mapB;  
 XX hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
 XX canine; veterinary; antibody; vaccine; immunisation.

XX Dermatophagoides pteronyssinus.

XX Key Location/Qualifiers

XX Modified-site 115..117 /note= "Asn is N-glycosylated"

XX Modified-site 240..242 /note= "Asn is N-glycosylated"

XX WO95449-A2.

XX 28-OCT-1999.

XX 16-APR-1999; 99WO-US008524.

XX 17-APR-1998; 98US-00062013.

XX 13-MAY-1998; 98US-0085235P.

XX 02-SEP-1998; 98US-0098909P.

XX (HESK-) HESKA CORP.

XX McCall CA, Hunter SM, Weber ER;

XX WPI; 2000-052700/04.

XX N-PSDB; AAZ38589; AAZ38590.

XX Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
 PT used to modify an animals' hypersensitivity to mite allergens.

XX Claim 3; Page 147-149; 154pp; English.

XX This sequence represents Dermatophagoides pteronyssinus mite allergen  
 CC protein (map) pDerp98-490, the mature form of pDerp98-509. pDerp98-490  
 CC has a molecular weight of 98 kD, comprising 490 amino acids, and has a  
 CC high degree of homology with the D. farinae mature 98 kD allergen, mapB  
 CC (AAU52525). Nucleic acid molecules encoding pDerp98-490 were isolated  
 CC from a D. pteronyssinus cDNA library by hybridisation with a probe  
 CC encoding the D. farinae high molecular weight map (HMM-map) composition.  
 CC Mite allergenic proteins and peptides, and nucleic acids encoding them,  
 CC may be used in therapeutic compositions to modify an animal's  
 CC hypersensitivity reaction to mite allergens. Animals that may be treated  
 CC include mammals and birds, especially felines, canines, equines, humans,  
 CC other pets, and work or domestic animals. The proteins or fragments may  
 CC also be used to diagnose allergies via a skin test. The proteins and  
 CC peptides can also be used to raise antibodies, which have a variety of

```
QY      1 DKLVMGVPEFYGRAXSIE 17  
        |||||  
Db     254 DKLVMGVPEFYGRAMSIE 270
```

RESULT 5	
AA52533	
ID	AA52533 standard; protein; 509 AA
XX	

DT	06-AUG-2003	(revised)
DT	22-FEB-2000	(first entry)

DE D. pteromyssus 98 kD mite allergen protein (map) Pderp98-509.  
XX  
XX Mite allergen protein; map; high molecular weight; HMW-map; allergy;  
KM house dust mite; IgE; immunoglobulin E; allergen; mapB;

	Key	Location/Qualification
XX	Dermatophagoides pteronyssinus	
OS		
XX		
FH		

FT	Peptide	
FT		1. 19
FT	/note= "Signal peptide"	
FT	20. 509	
FT	/note= "Mature Pderp98-509"	

PN W09954349-A2.

PD 28-OCT-1999

PF 16-APR-1999

PR 17-APR-1998; 98US-00062013.

PR 02-SEP-1998; 98US-0098909P.

PA (HESK-) HESKA CORP.

PI McCall CA, Hunter SW, Weber ER;

DR WPI; 2000-052700/04.

100

PT used to modify an animals' hypersensitivity to mite allergens

PS Claim 3; Page 134-136; 154pp; English.

CC This sequence represents Dermatophagoides pteronyssinus mite allergen  
CC protein (map). Pderp98-509 has a molecular weight of 98 kD,  
CC comprising 509 amino acids, and has a high degree of homology with the D  
CC faeine 98 kD allergen, map8 (A152523). Nucleic acid molecules encoding  
CC Pderp98-509 were isolated from a D. pteronyssinus cDNA library by  
CC hybridisation with a probe encoding the D. faeine high molecular weight  
CC map (HMM-map) composition. Mite allergenic proteins and peptides, and  
CC nucleic acids encoding them, may be used in therapeutic compositions to  
CC modify an animal's hypersensitivity reaction to mite allergens. Animals  
CC that may be treated include mammals and birds, especially felines, the  
CC canines, equines, humans, other pets, and work or domestic animals. The  
CC proteins or fragments may also be used to diagnose allergies via a skin  
CC test. The proteins and peptides can also be used to raise antibodies,  
CC which have a variety of potential uses. For example, they can be used as  
CC vaccines to passively immunise animals against dust mite

CC recover desired

XX  
SQ Sequence 509 AA;

Query Match 96.4%; Score 80; DB 3; Length 509;  
 Best Local Similarity 94.1%; Pred. No. 2.1e-06;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKLVMGVFPYGRAXSIE 17  
 |||||  
 DB 273 DKLVMGVFPYGRAXSIE 289

RESULT 6  
 AAU96337  
 ID AAU96337 standard; protein; 509 AA.

AC AAU96337;

DT 15-JUL-2002 (first entry)

DE Der HMW-map polypeptide #24.

XX Der HMW-map; American house dust mite; antiallergic; mite; IgE;  
 KW mite allergenic protein; immunoglobulin E; hypersensitivity;  
 KM immunocomplex formation.

OS Dermatophagoides farinae.

XX WO200222807-A2.

XX 21-MAR-2002.

PF 14-SEP-2001; 2001WO-US028730.

PR 14-SEP-2000; 2000US-00662293.

XX (HESK-) HESKA CORP.

PI Mccall CA, Hunter SW, Weber ER;

XX WPI; 2002-351888/38.

DR N-PSDB; ABK69581.

PT New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HMW-map protein, useful as a vaccine for treating mite allergy.

XX Claim 12; Page 134-136; 161pp; English.

CC The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic  
 CC acid. The Der HMW-map protein is useful for eliciting an immune response  
 CC against Der HMW-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitising a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig)E or Der HMW-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HMW-map polypeptides of the invention

XX Sequence 509 AA;

Query Match 96.4%; Score 80; DB 5; Length 509;  
 Best Local Similarity 94.1%; Pred. No. 2.1e-06;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKLVMGVFPYGRAXSIE 17  
 |||||  
 DB 273 DKLVMGVFPYGRAXSIE 289

RESULT 7

AAU96338  
 ID AAU96338 standard; protein; 509 AA.

AC AAU96338;

DT 15-JUL-2002 (first entry)

DE Der HMW-map polypeptide #25.

XX Der HMW-map; American house dust mite; antiallergic; mite; IgE;  
 KW mite allergenic protein; immunoglobulin E; hypersensitivity;  
 KM immunocomplex formation.

OS Dermatophagoides farinae.

XX WO200222807-A2.

XX 21-MAR-2002.

PF 14-SEP-2001; 2001WO-US028730.

PR 14-SEP-2000; 2000US-00662293.

XX (HESK-) HESKA CORP.

PI Mccall CA, Hunter SW, Weber ER;

XX WPI; 2002-351888/38.

DR N-PSDB; ABK69583.

PT New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HMW-map protein, useful as a vaccine for treating mite allergy.

XX Claim 12; Page 139-141; 161pp; English.

CC The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic  
 CC acid. The Der HMW-map protein is useful for eliciting an immune response  
 CC against Der HMW-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitising a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig)E or Der HMW-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HMW-map polypeptides of the invention

XX Sequence 509 AA;

Query Match 96.4%; Score 80; DB 5; Length 509;  
 Best Local Similarity 94.1%; Pred. No. 2.1e-06;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKLVMGVFPYGRAXSIE 17  
 |||||  
 DB 273 DKLVMGVFPYGRAXSIE 289

RESULT 8

AAU96338  
 ID AAU96338 standard; protein; 536 AA.

AC AAU96338;

DT 22-FEB-2000 (first entry)

DE House dust mite (D. farinae) mite allergen protein (map) PDerf98-536.

XX Mite allergen protein; map; high molecular weight; HMW-map; allergy;

KW house dust mite; IGE; immunoglobulin E; allergen; mapA; mapB;  
 KM hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
 KM canine; veterinary; antibody; vaccine; immunisation.  
 OS Dermatophagoides farinae.  
 XX WO9954349-A2.  
 PN 28-OCT-1999.  
 XX 16-APR-1999; 99WO-US008524.  
 PF 17-APR-1998; 98US-00062013.  
 PR 13-MAY-1998; 98US-0008295P.  
 PR 02-SEP-1998; 98US-0008909P.  
 PA (HESK-) HESKA CORP.  
 XX McCall CA, Hunter SW, Weber ER;  
 PI MPI; 2000-052700/04.  
 DR N-PSDB; AAZ38579, AAZ38580.  
 XX Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
 PT used to modify an animals' hypersensitivity to mite allergens.  
 PS Claim 3; Page 125-127; 154pp; English.  
 XX This sequence represents Dermatophagoides farinae mite allergen protein  
 CC (map) Pderf98-536, the mature form of Pderf98-555 (AAV52523). Pderf98-536  
 CC has a molecular weight of 98 kD, comprising 536 amino acids, and is a  
 CC component of the Dermatophagoides farinae high molecular weight mite  
 CC allergen protein (HMM-map) composition. The HMM-map composition was  
 CC isolated from a D. farinae homogenate by gel filtration, with each  
 CC fraction being analysed for the presence of proteins that bound to IGE  
 CC present in mite-allergic dog antisera. Mite allergenic proteins and  
 CC peptides, and nucleic acids encoding them, may be used in therapeutic  
 CC compositions to modify an animal's hypersensitivity reaction to mite  
 CC allergens. Animals that may be treated include mammals and birds,  
 CC especially felines, canines, equines, humans, other pets, and work or  
 CC domestic animals. The proteins or fragments may also be used to diagnose  
 CC allergies via a skin test. The proteins and peptides can also be used to  
 CC raise antibodies, which have a variety of potential uses. For example,  
 CC they can be used as vaccines to passively immunise animals against dust  
 CC mite hypersensitivity, as positive controls in test kits and as tools to  
 CC recover desired dust mite allergens from a mixture of proteins  
 XX SQ Sequence 536 AA;  
 Query Match 96.4%; Score 80; DB 3; Length 536;  
 Best Local Similarity 94.1%; Pred. No. 2.2e-06;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 DKLWGVFPYGRAXSIE 17  
 |||||  
 DB 254 DKLWGVFPYGRAXSIE 270  
 |||||  
 RESULT 9  
 AAU96329  
 ID AAU96329 standard; protein; 536 AA.  
 XX AC AAU96329;  
 XX 15-JUL-2002 (first entry)  
 DT Der HMM-map polypeptide #16.  
 XX Der HMM-map polypeptide #16.  
 KW Der HMM-map; American house dust mite; antiallergic; mite; IGE;  
 KM mite allergenic protein; immunoglobulin E; hypersensitivity;  
 KM immunocomplex formation.  
 XX OS Dermatophagoides farinae.

XX WO200222807-A2.  
 PN 21-MAR-2002.  
 XX 14-SEP-2001; 2001WO-US028730.  
 PF 14-SEP-2000; 2000US-00662293.  
 PR (HESK-) HESKA CORP.  
 XX McCall CA, Hunter SW, Weber ER;  
 PI MPI; 2002-351888/38.  
 DR N-PSDB; ABA69575.  
 XX New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HMM-map protein, useful as a vaccine for treating mite allergy.  
 PS Claim 12; Page 125-127; 161pp; English.  
 XX The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HMM-map protein, and its related nucleic  
 CC acid. The Der HMM-map protein is useful for eliciting an immune response  
 CC against Der HMM-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitising a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMM-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig)E or Der HMM-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HMM-map are useful for inhibiting  
 CC binding of proteins to IGE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HMM-map polypeptides of the invention  
 XX SQ Sequence 536 AA;  
 Query Match 96.4%; Score 80; DB 5; Length 536;  
 Best Local Similarity 94.1%; Pred. No. 2.2e-06;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 DKLWGVFPYGRAXSIE 17  
 |||||  
 DB 254 DKLWGVFPYGRAXSIE 270  
 |||||  
 RESULT 10  
 AAU52523  
 ID AAU52523 standard; protein; 555 AA.  
 XX AC AAU52523;  
 XX 22-FEB-2000 (first entry)  
 DT House dust mite (D. farinae) mite allergen protein (map) Pderf98-555.  
 XX Mite allergen protein; map; high molecular weight; HMM-map; allergy;  
 KM house dust mite; IGE; immunoglobulin E; allergen; mapA; mapB;  
 KM hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
 KM canine; veterinary; antibody; vaccine; immunisation.  
 XX OS Dermatophagoides farinae.  
 XX Key Location/Qualifiers  
 FH Peptide 1..19 "Signal peptide"  
 FT /note="555"  
 FT Protein 20..555  
 FT /note="Mature Pderf98-555"  
 XX PN WO9954349-A2.

PD 28-OCT-1999.  
 XX  
 PF 16-APR-1999; 99MO-US008524.  
 XX  
 PR 17-APR-1998; 98US-00062013.  
 PR 13-MAY-1998; 98US-0085295P.  
 PR 02-SEP-1998; 98US-0098909P.  
 XX  
 PA (HESK-) HESKA CORP.  
 XX  
 PI Mccall CA, Hunter SW, Weber ER;  
 XX  
 DR WPI; 2000-052700/04.  
 DR N-PSDB; AA238575, AA238576, AA238577, AA238578.  
 XX  
 PT Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
 XX used to modify an animals' hypersensitivity to mite allergens.  
 XX  
 PS Claim 3; Page 111-113; 154pp; English.  
 XX  
 CC This sequence represents Dermatophagoides farinae mite allergen protein  
 CC (map) Pderf38-555. Pderf38-555 has a molecular weight of 98 kD,  
 CC comprising 555 amino acids, and is a component of the Dermatophagoides  
 CC farinae high molecular weight mite allergen protein (Hmw-map).  
 CC composition. The Hmw-map composition was isolated from a D. farinae  
 CC homogenate by gel filtration, with each fraction being analysed for the  
 CC presence of proteins that bound to IgE present in mite-allergic dog  
 CC antisera. Mite allergenic proteins and peptides, and nucleic acids  
 CC encoding them, may be used in therapeutic compositions to modify an  
 CC animal's hypersensitivity reaction to mite allergens. Animals that may be  
 CC treated include mammals and birds, especially felines, canines, equines,  
 CC humans, other pets, and work or domestic animals. The proteins or  
 CC fragments may also be used to diagnose allergies via a skin test. The  
 CC proteins and peptides can also be used to raise antibodies, which have a  
 CC variety of potential uses. For example, they can be used as vaccines to  
 CC passively immunise animals against dust mite hypersensitivity, as  
 CC positive controls in test kits and as tools to recover desired dust mite  
 CC allergens from a mixture of proteins  
 XX  
 SQ Sequence 555 AA;  
 XX  
 Query Match 96.4%; Score 80; DB 3; Length 555;  
 Best Local Similarity 94.1%; Pred. No. 2.3e-06;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 DKLVNGVPEYGRAXSIE 17  
 DB 273 DKLVNGVPEYGRAXSIE 289  
 RESULT 11  
 AAU96327  
 ID AAU96327 standard; protein; 555 AA.  
 XX  
 AC AAU96327;  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX  
 DE Der Hmw-map polypeptide #14.  
 XX  
 XX Der Hmw-map; American house dust mite; anti-allergic; mite; IgE;  
 KM mite allergenic protein; immunoglobulin E; hypersensitivity;  
 KM immunocomplex formation.  
 XX  
 OS Dermatophagoides farinae.  
 XX  
 PN WO200222807-A2.  
 XX  
 PD 21-MAR-2002.  
 XX  
 PF 14-SEP-2001; 2001WO-US028730.  
 XX  
 PR 14-SEP-2000; 2000US-00662293.

XX  
 PA (HESK-) HESKA CORP.  
 XX  
 PI Mccall CA, Hunter SW, Weber ER;  
 XX  
 DR WPI; 2002-351888/38.  
 DR N-PSDB; ABK69571.  
 XX  
 PT New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der Hmw-map protein, useful as a vaccine for treating mite allergy.  
 XX  
 PS Claim 12; Page 114-116; 161pp; English.  
 XX  
 CC The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der Hmw-map protein, and its related nucleic  
 CC acid. The Der Hmw-map protein is useful for eliciting an immune response  
 CC against Der Hmw-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitising a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der Hmw-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig) E or Der Hmw-map protein activity associated with a  
 CC disease. Antibodies that bind to Der Hmw-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der Hmw-map polypeptides of the invention  
 XX  
 SQ Sequence 555 AA;  
 XX  
 Query Match 96.4%; Score 80; DB 5; Length 555;  
 Best Local Similarity 94.1%; Pred. No. 2.3e-06;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 DKLVNGVPEYGRAXSIE 17  
 DB 273 DKLVNGVPEYGRAXSIE 289  
 RESULT 12  
 AAU96328  
 ID AAU96328 standard; protein; 555 AA.  
 XX  
 AC AAU96328;  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX  
 DE Der Hmw-map polypeptide #15.  
 XX  
 XX Der Hmw-map; American house dust mite; anti-allergic; mite; IgE;  
 KM mite allergenic protein; immunoglobulin E; hypersensitivity;  
 KM immunocomplex formation.  
 XX  
 OS Dermatophagoides farinae.  
 XX  
 PN WO200222807-A2.  
 XX  
 PD 21-MAR-2002.  
 XX  
 PF 14-SEP-2001; 2001WO-US028730.  
 XX  
 PR 14-SEP-2000; 2000US-00662293.  
 XX  
 PA (HESK-) HESKA CORP.  
 XX  
 PI Mccall CA, Hunter SW, Weber ER;  
 XX  
 DR WPI; 2002-351888/38.  
 DR N-PSDB; ABK69573.  
 XX  
 PT New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der Hmw-map protein, useful as a vaccine for treating mite allergy.

XX Claim 12; Page 120-122; 16pp; English.  
PS  
XX  
CC The invention relates to an isolated mite allergenic protein of  
CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic  
CC acid. The Der HMW-map protein is useful for eliciting an immune response  
CC against Der HMW-map protein. The protein or a reagent comprising a non-  
CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
CC cat) susceptible to or having an allergic response to a mite. A  
CC therapeutic composition is useful for desensitizing a host animal to an  
CC allergic response to a mite. The DNA and protein can be used in the  
CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
CC of immunoglobulin (Ig) E or Der HMW-map protein activity associated with a  
CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
CC reducing hypersensitivity responses to mite allergens, and as vaccines  
CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
CC represent Der HMW-map polypeptides of the invention  
XX  
SQ Sequence 555 AA;  
XX  
Query Match 96.4%; Score 80; DB 5; Length 555;  
Best Local Similarity 94.1%; Pred. No. 2.3e-06;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 DKLVGVPFYGRAXSI 17  
DB 273 DKLVGVPFYGRAXSI 289  
XX  
RESULT 13  
ABR64366  
ID ABR64366 standard; protein; 460 AA.  
XX  
XX ABR64366;  
AC  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 19890.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical.  
XX  
XX Drosophila melanogaster.  
OS  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
XX  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PMD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
XX  
DR N-PSDB; ABL08469.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signaling and cell-cell  
XX interactions.  
XX  
PS Disclosure; SEQ ID NO 19890; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signaling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins (ABR57737-  
CC ABR72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pat\_sequences  
XX  
SQ Sequence 460 AA;  
XX  
Query Match 72.3%; Score 60; DB 4; Length 460;  
Best Local Similarity 56.2%; Pred. No. 0.012;  
Matches 9; Conservative 6; Mismatches 1; Indels 0; Gaps 0;  
OY 1 DKLVGVPFYGRAXSI 16  
DB 255 EKILVGFYGRSFTL 270  
XX  
RESULT 14  
AAE28202  
ID AAE28202 standard; protein; 489 AA.  
XX  
AC AAE28202;  
XX  
DT 27-DEC-2002 (first entry)  
XX  
DE Flea chitinase protein, PCFCHT489.  
XX  
XX Flea; chitinase; enzyme; flea infestation; gene therapy; insecticide;  
XX CHT.  
XX  
XX Ctenocephalides felis.  
XX  
PN US6416977-B1.  
XX  
PD 09-JUL-2002.  
XX  
PF 07-APR-2000; 2000US-00545814.  
XX  
PR 09-APR-1999; 99US-0128833P.  
XX  
PA (HESK-) HESKA CORP.  
XX  
PI Becher AM;  
XX  
DR WPI; 2002-634733/68.  
XX  
DR N-PSDB; AAD45161.  
XX  
PT Novel isolated flea chitinase nucleic acid useful for protecting an  
XX animal from flea infestation or to reduce flea infestation in an animal  
XX susceptible to flea infestation.  
XX  
PS Example 1; Col 81-84; 50pp; English.  
XX  
XX The present invention relates to flea chitinase (CHT) proteins and their  
XX corresponding polynucleotides. Sequences of the invention are useful for  
XX protecting an animal from flea infestation or to reduce flea infestation  
XX in an animal susceptible to flea infestation. They are also used in gene  
XX therapy. The present sequence is flea chitinase protein  
XX  
SQ Sequence 489 AA;  
XX  
Query Match 71.1%; Score 59; DB 5; Length 489;  
Best Local Similarity 62.5%; Pred. No. 0.02;  
Matches 10; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
OY 1 DKLVGVPFYGRAXSI 16  
DB 199 NKLWGVFPYGRSFTL 214  
XX  
RESULT 15  
AAE28199  
ID AAE28199 standard; protein; 559 AA.  
XX

AC AAE28199 ;

DT 27-DEC-2002 (first entry)

DE Flea PCfCHT583 mature chitinase protein, PCfCHT559.

KM Flea; chitinase; enzyme; flea infestation; gene therapy; insecticide  
KM  
KM CHT.

*Ctenocephalides felis*.

PN - US6416977-B1.

PD 09-JUL-2002.  
xx

PF 07-APR-2000; 2000US-00545814.

PR 09-APR-1999; 99US-0128833P.

PA (HESK-) HESKA CORP.

PI Becher AM;

DR WPI; 2002-634733/68.  
DR N-PSDB; AAD45152.

PT Novel isolated flea chitinase nucleic acid useful for protecting an animal from flea infestation or to reduce flea infestation in an animal susceptible to flea infestation.

PS Claim 4; Col 69-72; 50pp; English.

CC The present invention relates to flea chitinase (ChT) proteins and their  
CC corresponding polynucleotides. Sequences of the invention are useful for  
CC protecting an animal from flea infestation or to reduce flea infestation  
CC in an animal susceptible to flea infestation. They are also used in gene  
CC therapy. The present sequence is flea mature chitinase protein

Sequence 559 AA;

Query Match	Score	DB	Length
71.1%	59	5	559

Best Local Similarity: 62.5%; Pred. No. 0.023;  
Matches: 10; Conservative: 5; Mismatches: 1; Indels: 0; Gaps: 0;

QY 1 DKLVMEVPFYGRAXSI 16

Db 242 NKLWGVFPFYGRSFTL 257

Search completed: March 22, 2004, 06:51:42  
Job time : 6.16898 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:59:34 ; Search time 3.47749 Seconds

(without alignments)  
1265.926 Million cell updates/sec

Title: US-09-662-293-10

Sequence: 1 DKLWGVFPRYGRAXSIE 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1049977 seqs, 258955339 residues

Total number of hits satisfying chosen parameters: 1049977

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	81	97.6	17	US-10-218-743-10	Sequence 10, Appl
2	80	96.4	490	US-10-218-743-41	Sequence 41, Appl
3	80	96.4	509	US-10-218-743-35	Sequence 35, Appl
4	80	96.4	509	US-10-218-743-38	Sequence 38, Appl
5	80	96.4	536	US-10-218-743-21	Sequence 21, Appl
6	80	96.4	555	US-10-218-743-11	Sequence 15, Appl
7	80	96.4	555	US-10-218-743-18	Sequence 18, Appl
8	62	74.7	428	US-10-369-493-16644	Sequence 16644, A
9	66.3	449	15	US-10-369-493-3308	Sequence 3308, Ap
10	66.3	572	15	US-10-369-493-18585	Sequence 18585, A
11	66.3	559	15	US-10-369-493-17162	Sequence 17162, A
12	63.4	392	15	US-10-369-493-19361	Sequence 19361, A
13	61.4	678	15	US-10-295-027-632	Sequence 632, App
14	60.2	383	9	US-09-459-749D-17	Sequence 17, Appl

16	50	60.2	407	15	US-10-369-493-20271	Sequence 20271, A
17	49	60.2	485	15	US-10-369-493-3399	Sequence 3399, Ap
18	50	59.0	216	15	US-10-425-114-70064	Sequence 70064, A
19	49	59.0	383	14	US-10-097-340-45	Sequence 45, Appl
20	49	59.0	383	15	US-10-295-027-270	Sequence 270, App
21	49	59.0	399	14	US-10-316-754-11	Sequence 11, Appl
22	49	59.0	399	14	US-10-316-754-19	Sequence 19, Appl
23	48	57.8	170	14	US-10-218-743-44	Sequence 44, Appl
24	48	57.8	462	15	US-10-369-493-8244	Sequence 8244, Ap
25	47	56.6	373	14	US-10-161-547-14	Sequence 14, Appl
26	47	56.6	373	14	US-10-161-547-15	Sequence 15, Appl
27	47	56.6	385	15	US-10-295-027-278	Sequence 278, App
28	47	56.6	423	15	US-10-373-802-2	Sequence 2, Appl1
29	47	56.6	445	14	US-10-004-219B-10	Sequence 10, Appl1
30	47	56.6	445	15	US-10-369-493-3089	Sequence 3089, Ap
31	47	56.6	457	15	US-10-094-749-1900	Sequence 1900, Ap
32	47	56.6	466	14	US-10-161-547-2	Sequence 2, Appl1
33	47	56.6	466	14	US-10-161-547-4	Sequence 4, Appl1
34	46	55.4	234	12	US-10-425-114-51995	Sequence 51995, A
35	46	55.4	314	12	US-10-425-114-57688	Sequence 57688, A
36	46	55.4	316	12	US-10-425-114-57443	Sequence 57443, A
37	45	54.2	371	9	US-09-923-844B-2	Sequence 2, Appl1
38	44	53.0	546	15	US-10-369-493-18584	Sequence 18584, A
39	44	53.0	617	15	US-10-369-493-6743	Sequence 6743, Ap
40	44	51.8	180	12	US-10-424-599-26823	Sequence 26823, A
41	43	51.8	358	12	US-10-425-114-47084	Sequence 47084, A
42	43	51.8	365	12	US-10-424-599-273219	Sequence 273219, A
43	43	51.8	376	9	US-09-748-033-3	Sequence 3, Appl1
44	43	51.8	427	14	US-10-156-761-14446	Sequence 14446, A
45	43	51.8	1686	15	US-10-369-493-13058	Sequence 13058, A

## ALIGNMENTS

RESULT 1  
US-10-218-743-10  
; Sequence 10, Application US/10218743  
; Publication No. US20030096779A1  
; GENERAL INFORMATION:  
; APPLICANT: McCall, Catherine A.  
; APPLICANT: Hunter, Shirley Wu  
; APPLICANT: Weber, Eric R.  
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: AL-2-C3  
; CURRENT APPLICATION NUMBER: US/10/218,743  
; CURRENT FILING DATE: 2002-08-13  
; PRIOR APPLICATION NUMBER: US/09/292,225  
; PRIOR FILING DATE: 1999-04-15  
; PRIOR APPLICATION NUMBER: 60/098,909  
; PRIOR FILING DATE: 1998-09-02  
; PRIOR APPLICATION NUMBER: 60/085,295  
; PRIOR FILING DATE: 1998-05-13  
; PRIOR APPLICATION NUMBER: 60/098,565  
; PRIOR FILING DATE: 1998-04-17  
; PRIOR APPLICATION NUMBER: 09/062,013  
; PRIOR FILING DATE: 1998-04-17  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 17  
; TYPE: PRT  
; ORGANISM: Dermatophagoides farinae  
; FEATURE:  
; OTHER INFORMATION: At location 14, Xaa = any amino acid  
US-10-218-743-10  
Query Match 97.6%; Score 81; DB 14; Length 17;  
Best Local Similarity 100.0%; Pred. No. 8.7e-08;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 DKLWGVFPRYGRAXSIE 17

Db 1 DKLWGVFPYGRAXSIE 17

RESULT 2  
US-10-218-743-41  
Sequence 41, Application US/10218743  
Publication No. US20030096779A1  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley Wu  
APPLICANT: Weber, Eric R.  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
TITLE OF INVENTION: AND USES THEREOF  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/10/218,743  
CURRENT FILING DATE: 2002-08-13  
PRIOR APPLICATION NUMBER: US/09/292,225  
PRIOR FILING DATE: 1999-04-15  
PRIOR APPLICATION NUMBER: 60/098,909  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/085,295  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/098,565  
PRIOR FILING DATE: 1998-04-17  
PRIOR APPLICATION NUMBER: 09/062,013  
PRIOR FILING DATE: 1998-04-17  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 41  
LENGTH: 490  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-10-218-743-41

Query Match 96.4%; Score 80; DB 14; Length 490;  
Best Local Similarity 94.1%; Pred. No. 5.2e-06;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Cy 1 DKLWGVFPYGRAXSIE 17  
Db 254 DKLWGVFPYGRAXSIE 270

RESULT 3  
US-10-218-743-35  
Sequence 35, Application US/10218743  
Publication No. US20030096779A1  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley Wu  
APPLICANT: Weber, Eric R.  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
TITLE OF INVENTION: AND USES THEREOF  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/10/218,743  
CURRENT FILING DATE: 2002-08-13  
PRIOR APPLICATION NUMBER: US/09/292,225  
PRIOR FILING DATE: 1999-04-15  
PRIOR APPLICATION NUMBER: 60/098,909  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/085,295  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/098,565  
PRIOR FILING DATE: 1998-04-17  
PRIOR APPLICATION NUMBER: 09/062,013  
PRIOR FILING DATE: 1998-04-17  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 35  
LENGTH: 509  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae

US-10-218-743-35

Query Match 96.4%; Score 80; DB 14; Length 509;  
Best Local Similarity 94.1%; Pred. No. 5.4e-06;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Cy 1 DKLWGVFPYGRAXSIE 17  
Db 273 DKLWGVFPYGRAXSIE 289

RESULT 4  
US-10-218-743-38  
Sequence 38, Application US/10218743  
Publication No. US20030096779A1  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley Wu  
APPLICANT: Weber, Eric R.  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
TITLE OF INVENTION: AND USES THEREOF  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/10/218,743  
CURRENT FILING DATE: 2002-08-13  
PRIOR APPLICATION NUMBER: US/09/292,225  
PRIOR FILING DATE: 1999-04-15  
PRIOR APPLICATION NUMBER: 60/098,909  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/085,295  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/098,565  
PRIOR FILING DATE: 1998-04-17  
PRIOR APPLICATION NUMBER: 09/062,013  
PRIOR FILING DATE: 1998-04-17  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 38  
LENGTH: 509  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-10-218-743-38

Query Match 96.4%; Score 80; DB 14; Length 509;  
Best Local Similarity 94.1%; Pred. No. 5.4e-06;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Cy 1 DKLWGVFPYGRAXSIE 17  
Db 273 DKLWGVFPYGRAXSIE 289

RESULT 5  
US-10-218-743-21  
Sequence 21, Application US/10218743  
Publication No. US20030096779A1  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley Wu  
APPLICANT: Weber, Eric R.  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
TITLE OF INVENTION: AND USES THEREOF  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/10/218,743  
CURRENT FILING DATE: 2002-08-13  
PRIOR APPLICATION NUMBER: US/09/292,225  
PRIOR FILING DATE: 1999-04-15  
PRIOR APPLICATION NUMBER: 60/098,909  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/085,295  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/098,565  
PRIOR FILING DATE: 1998-04-17  
PRIOR APPLICATION NUMBER: 09/062,013

;; PRIOR FILING DATE: 1998-04-17  
;; NUMBER OF SEQ ID NOS: 49  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 21  
;; LENGTH: 536  
;; TYPE: PRT  
;; ORGANISM: Dermatophagoides farinae  
US-10-218-743-21

Query Match 96.4%; Score 80; DB 14; Length 536;  
Best Local Similarity 94.1%; Pred. No. 5.9e-06;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKLWGVFPFYGRAXSIE 17  
|||  
DB 254 DKLWGVFPFYGRAXSIE 270

RESULT 6  
US-10-218-743-15

;; Sequence 15, Application US/10218743  
;; Publication No. US20030096779A1

;; GENERAL INFORMATION:

;; APPLICANT: McCall, Catherine A.

;; APPLICANT: Hunter, Shirley Wu

;; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS

;; FILE REFERENCE: AL-2-C3

;; CURRENT APPLICATION NUMBER: US/10/218,743

;; PRIOR FILING DATE: 2002-08-13

;; PRIOR APPLICATION NUMBER: US/09/292,225

;; PRIOR FILING DATE: 1999-04-15

;; PRIOR APPLICATION NUMBER: 60/098,909

;; PRIOR FILING DATE: 1998-09-02

;; PRIOR APPLICATION NUMBER: 60/085,295

;; PRIOR FILING DATE: 1998-05-13

;; PRIOR APPLICATION NUMBER: 60/098,565

;; PRIOR FILING DATE: 1998-04-17

;; PRIOR APPLICATION NUMBER: 09/062,013

;; PRIOR FILING DATE: 1998-04-17

;; NUMBER OF SEQ ID NOS: 49

;; SOFTWARE: PatentIn Ver. 2.0

;; SEQ ID NO 15

;; LENGTH: 555

;; TYPE: PRT

;; ORGANISM: Dermatophagoides farinae

US-10-218-743-15

Query Match 96.4%; Score 80; DB 14; Length 555;  
Best Local Similarity 94.1%; Pred. No. 5.9e-06;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKLWGVFPFYGRAXSIE 17  
|||  
DB 273 DKLWGVFPFYGRAXSIE 289

;; PRIOR APPLICATION NUMBER: 60/098,909  
;; PRIOR FILING DATE: 1998-09-02  
;; PRIOR APPLICATION NUMBER: 60/085,295  
;; PRIOR FILING DATE: 1998-05-13  
;; PRIOR APPLICATION NUMBER: 60/098,565  
;; PRIOR FILING DATE: 1998-04-17  
;; PRIOR APPLICATION NUMBER: 09/062,013  
;; PRIOR FILING DATE: 1998-04-17  
;; NUMBER OF SEQ ID NOS: 49  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 18  
;; LENGTH: 555  
;; TYPE: PRT  
;; ORGANISM: Dermatophagoides farinae  
US-10-218-743-18

Query Match 96.4%; Score 80; DB 14; Length 555;  
Best Local Similarity 94.1%; Pred. No. 5.9e-06;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKLWGVFPFYGRAXSIE 17  
|||  
DB 273 DKLWGVFPFYGRAXSIE 289

RESULT 8

US-10-369-493-16644

;; Sequence 16644, Application US/10369493  
;; Publication No. US20030233675A1

;; GENERAL INFORMATION:

;; APPLICANT: Cao, Yongwei

;; APPLICANT: Hinkle, Gregory J.

;; APPLICANT: Slater, Steven C.

;; APPLICANT: Goldstein, Barry S.

;; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

;; FILE REFERENCE: 38-10(52052)B

;; CURRENT APPLICATION NUMBER: US/10/369,493

;; PRIOR FILING DATE: 2003-02-28

;; PRIOR APPLICATION NUMBER: US 60/360,039

;; PRIOR FILING DATE: 2002-02-21

;; NUMBER OF SEQ ID NOS: 47374

;; SEQ ID NO 16644

;; LENGTH: 428

;; TYPE: PRT

;; ORGANISM: Bacillus thuringiensis

US-10-369-493-16644

Query Match 74.7%; Score 62; DB 15; Length 428;  
Best Local Similarity 91.7%; Pred. No. 0.0083;  
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DKLWGVFPFYGR 12  
|||  
DB 178 DKLWGVFPFYGR 189

RESULT 9

US-10-424-599-209667

;; Sequence 209667, Application US/10424599  
;; Publication No. US20040031072A1

;; GENERAL INFORMATION:

;; APPLICANT: La Rosa, Thomas J

;; APPLICANT: Kowalczyk, David K

;; APPLICANT: Zhou, Yihua

;; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

;; FILE REFERENCE: 38-21(53223)B

;; CURRENT APPLICATION NUMBER: US/10/424,599

;; PRIOR FILING DATE: 2003-04-28

;; NUMBER OF SEQ ID NOS: 285684

```

; SEQ ID NO 209667
; LENGTH: 366
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_31357C.1.pap
US-10-424-599-209667

```

```

Query Match          66.3%; Score 55; DB 12; Length 366;
Best Local Similarity 66.7%; Pred. No. 0.13;
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      2 KLVMGVFPYGRAXSI 16
Db      249 KLVMGGLPIYGRAMAL 263

```

```

RESULT 10
US-10-369-493-3308
; Sequence 3308, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3308
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Neurospora crassa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(449)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-3308

```

```

Query Match          66.3%; Score 55; DB 15; Length 449;
Best Local Similarity 69.2%; Pred. No. 0.16;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 DKLVGVFPYGRGA 13
Db      307 DKVLGMFLYGRA 319

```

```

RESULT 11
US-10-369-493-18585
; Sequence 18585, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 18585

```

```

; LENGTH: 572
; TYPE: PRT
; ORGANISM: Halobacterium sp. NRC-1
US-10-369-493-18585

```

```

Query Match          66.3%; Score 55; DB 15; Length 572;
Best Local Similarity 60.0%; Pred. No. 0.21;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      3 LVMGVFPYGRAXSIE 17
Db      433 LVLMGFYGRGCVNE 447

```

```

RESULT 12
US-10-369-493-17162
; Sequence 17162, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 17162
; LENGTH: 599
; TYPE: PRT
; ORGANISM: Bacillus halodurans
US-10-369-493-17162

```

```

Query Match          66.3%; Score 55; DB 15; Length 599;
Best Local Similarity 71.4%; Pred. No. 0.22;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      2 KLVMGVFPYGRAXS 15
Db      325 KLVLMGFYGRGWS 338

```

```

RESULT 13
US-10-369-493-19361
; Sequence 19361, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 19361
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-10-369-493-19361

```

```

Query Match          63.9%; Score 53; DB 15; Length 392;
Best Local Similarity 81.8%; Pred. No. 0.33;

```

Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 KLVGVFPYGR 12

Db 267 KLVGVFPYGR 277

## RESULT 14

US-10-295-027-632

; Sequence 632, Application US/10295027

; Publication No. US2003023250A1

; GENERAL INFORMATION:

; APPLICANT: Afar, Daniel

; APPLICANT: Aziz, Natasha

; APPLICANT: Ginsberg, Wendy M.

; APPLICANT: Gish, Kurt C.

; APPLICANT: Glynn, Richard

; APPLICANT: Hevez, Peter A.

; APPLICANT: Mack, David H.

; APPLICANT: Murray, Richard

; APPLICANT: Watson, Susan R.

; APPLICANT: Eos Biotechnology, Inc.

; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and

; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer

; FILE REFERENCE: 018501-012500US

; CURRENT APPLICATION NUMBER: US/10/295,027

; CURRENT FILING DATE: 2002-11-13

; PRIOR APPLICATION NUMBER: US 09/663,733

; PRIOR FILING DATE: 2000-09-15

; PRIOR APPLICATION NUMBER: US 60/350,666

; PRIOR FILING DATE: 2001-11-13

; PRIOR APPLICATION NUMBER: US 60/335,394

; PRIOR FILING DATE: 2001-11-15

; PRIOR APPLICATION NUMBER: US 60/332,464

; PRIOR FILING DATE: 2001-11-21

; PRIOR APPLICATION NUMBER: US 60/334,393

; PRIOR FILING DATE: 2001-11-29

; PRIOR APPLICATION NUMBER: US 60/340,376

; PRIOR FILING DATE: 2001-12-14

; PRIOR APPLICATION NUMBER: US 60/347,211

; PRIOR FILING DATE: 2002-01-08

; PRIOR APPLICATION NUMBER: US 60/347,349

; PRIOR FILING DATE: 2002-01-10

; PRIOR APPLICATION NUMBER: US 60/355,250

; PRIOR FILING DATE: 2002-02-08

; PRIOR APPLICATION NUMBER: US 60/356,714

; PRIOR FILING DATE: 2002-02-13

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 1386

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 632

; LENGTH: 678

; TYPE: PRT

; ORGANISM: Homo sapiens

; US-10-295-027-632

Query Match 61.4%; Score 51; DB 15; Length 678;

Best Local Similarity 66.7%; Pred. No. 1.4; Indels 0; Gaps 0;

Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKLWGVFPYGR 12

Db 254 KLVGVFPYGR 265

## RESULT 15

US-09-459-749D-17

; Sequence 17, Application US/09459749D

; Patent No. US20020136716A1

; GENERAL INFORMATION:

; APPLICANT: Millis, Albert J. T.

; TITLE OF INVENTION: Compositions and Methods for Altering Cell Migration

; FILE REFERENCE: 0794.016A

; CURRENT APPLICATION NUMBER: US/09/459,749D

; CURRENT FILING DATE: 1999-12-10

; PRIOR APPLICATION NUMBER: 60/111,856

; PRIOR FILING DATE: 1998-12-11

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 17

; LENGTH: 383

; TYPE: PRT

; ORGANISM: Sus scrofa

; US-09-459-749D-17

Query Match 60.2%; Score 50; DB 9; Length 383;

Best Local Similarity 50.0%; Pred. No. 1.1; Indels 2; Gaps 0;

Matches 8; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 1 DKLWGVFPYGRAXSI 16

Db 252 NKLWGVFPYGRSFTL 267

Search completed: March 22, 2004, 07:45:49  
Job time : 3.47749 secs



```
FILE REFERENCE: AL-2-C3
CURRENT APPLICATION NUMBER: US/09/292,225
CURRENT FILING DATE: 1999-04-15
EARLIER APPLICATION NUMBER: 60/098,909
EARLIER FILING DATE: 1998-09-02
EARLIER APPLICATION NUMBER: 60/085,295
EARLIER FILING DATE: 1998-05-13
EARLIER APPLICATION NUMBER: 60/098,565
EARLIER FILING DATE: 1998-04-17
EARLIER APPLICATION NUMBER: 09/062,013
EARLIER FILING DATE: 1998-04-17
NUMBER OF SEQ ID NOS: 49
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 41
LENGTH: 490
TYPE: PRT
ORGANISM: Dermatophagoides farinae
US-09-292-225-41
```

```
Query Match          96.4%; Score 80; DB 4; Length 490;
Best Local Similarity 94.1%; Pred. No. 7.8e-07;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY      1 DKLWGVFPYGRAXSIE 17
      |||||
Db      254 DKLWGVFPYGRAXSIE 270
```

```
RESULT 3
US-09-292-225-35
Sequence 35, Application US/09292225
Patent No. 6455686
```

```
GENERAL INFORMATION:
APPLICANT: McCall, Catherine A.
APPLICANT: Hunter, Shirley Wu
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
FILE REFERENCE: AL-2-C3
CURRENT APPLICATION NUMBER: US/09/292,225
CURRENT FILING DATE: 1999-04-15
EARLIER APPLICATION NUMBER: 60/098,909
EARLIER FILING DATE: 1998-09-02
EARLIER APPLICATION NUMBER: 60/085,295
EARLIER FILING DATE: 1998-05-13
EARLIER APPLICATION NUMBER: 60/098,565
EARLIER FILING DATE: 1998-04-17
EARLIER APPLICATION NUMBER: 09/062,013
EARLIER FILING DATE: 1998-04-17
NUMBER OF SEQ ID NOS: 49
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 35
LENGTH: 509
TYPE: PRT
ORGANISM: Dermatophagoides farinae
US-09-292-225-35
```

```
Query Match          96.4%; Score 80; DB 4; Length 509;
Best Local Similarity 94.1%; Pred. No. 8.1e-07;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY      1 DKLWGVFPYGRAXSIE 17
      |||||
Db      273 DKLWGVFPYGRAXSIE 289
```

```
RESULT 4
US-09-292-225-38
Sequence 38, Application US/09292225
Patent No. 6455686
GENERAL INFORMATION:
APPLICANT: McCall, Catherine A.
APPLICANT: Hunter, Shirley Wu
```

```
APPLICANT: Weber, Eric R.
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
FILE REFERENCE: AL-2-C3
CURRENT APPLICATION NUMBER: US/09/292,225
CURRENT FILING DATE: 1999-04-15
EARLIER APPLICATION NUMBER: 60/098,909
EARLIER FILING DATE: 1998-09-02
EARLIER APPLICATION NUMBER: 60/085,295
EARLIER FILING DATE: 1998-05-13
EARLIER APPLICATION NUMBER: 60/098,565
EARLIER FILING DATE: 1998-04-17
EARLIER APPLICATION NUMBER: 09/062,013
EARLIER FILING DATE: 1998-04-17
NUMBER OF SEQ ID NOS: 49
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 38
LENGTH: 509
TYPE: PRT
ORGANISM: Dermatophagoides farinae
US-09-292-225-38
```

```
Query Match          96.4%; Score 80; DB 4; Length 509;
Best Local Similarity 94.1%; Pred. No. 8.1e-07;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY      1 DKLWGVFPYGRAXSIE 17
      |||||
Db      273 DKLWGVFPYGRAXSIE 289
```

```
RESULT 5
US-09-292-225-21
Sequence 21, Application US/09292225
Patent No. 6455686
GENERAL INFORMATION:
APPLICANT: McCall, Catherine A.
APPLICANT: Hunter, Shirley Wu
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
FILE REFERENCE: AL-2-C3
CURRENT APPLICATION NUMBER: US/09/292,225
CURRENT FILING DATE: 1999-04-15
EARLIER APPLICATION NUMBER: 60/098,909
EARLIER FILING DATE: 1998-09-02
EARLIER APPLICATION NUMBER: 60/085,295
EARLIER FILING DATE: 1998-05-13
EARLIER APPLICATION NUMBER: 60/098,565
EARLIER FILING DATE: 1998-04-17
EARLIER APPLICATION NUMBER: 09/062,013
EARLIER FILING DATE: 1998-04-17
NUMBER OF SEQ ID NOS: 49
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 21
LENGTH: 536
TYPE: PRT
ORGANISM: Dermatophagoides farinae
US-09-292-225-21
```

```
Query Match          96.4%; Score 80; DB 4; Length 536;
Best Local Similarity 94.1%; Pred. No. 8.6e-07;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY      1 DKLWGVFPYGRAXSIE 17
      |||||
Db      254 DKLWGVFPYGRAXSIE 270
```

```
RESULT 6
US-09-292-225-15
Sequence 15, Application US/09292225
Patent No. 6455686
```

```

; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; EARLIER FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: 60/098,909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085,295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098,565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062,013
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 15
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-09-292-225-15
```

```

Query Match          96.4%; Score 80; DB 4; Length 555;
Best Local Similarity 94.1%; Pred. No. 9e-07;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```

QY 1 DKLWGVFPFYGRAXSI 17
    |||||
Db 273 DKLWGVFPFYGRAXSI 289
```

```

RESULT 7
US-09-292-225-18
; Sequence 18, Application US/09292225
; Patent No. 6455686
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; EARLIER FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: 60/098,909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085,295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098,565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062,013
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 18
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-09-292-225-18
```

```

Query Match          96.4%; Score 80; DB 4; Length 555;
Best Local Similarity 94.1%; Pred. No. 9e-07;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```

QY 1 DKLWGVFPFYGRAXSI 17
    |||||
Db 273 DKLWGVFPFYGRAXSI 289
```

RESULT 8

```

US-09-545-814-29
; Sequence 29, Application US/09545814
; Patent No. 6416977
; GENERAL INFORMATION:
; APPLICANT: Becher, Anna M.
; TITLE OF INVENTION: FLEA CHITINASE NUCLEIC ACID MOLECULES, PROTEINS AND
; FILE REFERENCE: FC-5-C1
; CURRENT APPLICATION NUMBER: US/09/545,814
; EARLIER FILING DATE: 2000-04-07
; EARLIER APPLICATION NUMBER: 60/128,833
; EARLIER FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 29
; LENGTH: 489
; TYPE: PRT
; ORGANISM: Tagged Ctenocephalides felis
US-09-545-814-29
```

```

Query Match          71.1%; Score 59; DB 4; Length 489;
Best Local Similarity 62.5%; Pred. No. 0.0068;
Matches 10; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
```

```

QY 1 DKLWGVFPFYGRAXSI 16
    :|||:|||||:
Db 199 NKLWGVFPFYGRSFTL 214
```

```

RESULT 9
US-09-545-814-14
; Sequence 14, Application US/09545814
; Patent No. 6416977
; GENERAL INFORMATION:
; APPLICANT: Becher, Anna M.
; TITLE OF INVENTION: FLEA CHITINASE NUCLEIC ACID MOLECULES, PROTEINS AND
; FILE REFERENCE: FC-5-C1
; CURRENT APPLICATION NUMBER: US/09/545,814
; EARLIER FILING DATE: 2000-04-07
; EARLIER APPLICATION NUMBER: 60/128,833
; EARLIER FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 14
; LENGTH: 559
; TYPE: PRT
; ORGANISM: Ctenocephalides felis
US-09-545-814-14
```

```

Query Match          71.1%; Score 59; DB 4; Length 559;
Best Local Similarity 62.5%; Pred. No. 0.008;
Matches 10; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
```

```

QY 1 DKLWGVFPFYGRAXSI 16
    :|||:|||||:
Db 242 NKLWGVFPFYGRSFTL 257
```

```

RESULT 10
US-09-545-814-2
; Sequence 2, Application US/09545814
; Patent No. 6416977
; GENERAL INFORMATION:
; APPLICANT: Becher, Anna M.
; TITLE OF INVENTION: FLEA CHITINASE NUCLEIC ACID MOLECULES, PROTEINS AND
; FILE REFERENCE: FC-5-C1
; CURRENT APPLICATION NUMBER: US/09/545,814
; EARLIER FILING DATE: 2000-04-07
; EARLIER APPLICATION NUMBER: 60/128,833
; EARLIER FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 40
```



SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 583  
TYPE: PRT  
ORGANISM: Ctenocephalides felis  
US-09-545-814-2

Query Match  
Best Local Similarity 71.1%; Score 59; DB 4; Length 583;  
Best Local Similarity 62.5%; Pred. No. 0.0084;  
Matches 10; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKLWGVPPYGRAXSI 16  
:|||||: :  
Db 266 NKLWVGIPYGRSFTL 281

RESULT 11  
US-09-545-814-5  
Sequence 5, Application US/09545814  
Patent No. 6416977  
GENERAL INFORMATION:  
APPLICANT: Becher, Anna M.  
TITLE OF INVENTION: FLEA CHITINASE NUCLEIC ACID MOLECULES, PROTEINS AND  
FILE REFERENCE: PC-5-C1  
CURRENT APPLICATION NUMBER: US/09/545,814  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: 60/128,833  
NUMBER OF SEQ ID NOS: 40  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5  
LENGTH: 583  
TYPE: PRT  
ORGANISM: Ctenocephalides felis  
US-09-545-814-5

Query Match  
Best Local Similarity 71.1%; Score 59; DB 4; Length 583;  
Best Local Similarity 62.5%; Pred. No. 0.0084;  
Matches 10; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKLWGVPPYGRAXSI 16  
:|||||: :  
Db 266 NKLWVGIPYGRSFTL 281

RESULT 12  
US-08-524-051-2  
Sequence 2, Application US/08524051  
Patent No. 5866788  
GENERAL INFORMATION:  
APPLICANT: Kramer, Karl J.  
APPLICANT: Muthukrishnan, Subbaratnam  
APPLICANT: Choi, Hee Kyung  
APPLICANT: Corpuz, Lolita  
APPLICANT: Gopalakrishnan, Bhuvana  
TITLE OF INVENTION: RECOMBINANT CHITINASE AND USE THEREOF AS  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hovey, Williams, Timmons & Collins  
STREET: 2405 Grand Blvd., Suite 400  
CITY: Kansas City  
STATE: MO  
COUNTRY: USA  
ZIP: 64108  
COMPUTER READABLE FORM:  
MEDIUM TYPE: IBM PC compatible  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/524,051

FILING DATE: 800  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Collins, John M.  
REGISTRATION NUMBER: 26,262  
REFERENCE/DOCKET NUMBER: 22875-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (816)474-9050  
TELEFAX: (816)474-9057  
TELEX: 434-363  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 554 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-524-051-2

Query Match  
Best Local Similarity 69.9%; Score 58; DB 2; Length 554;  
Best Local Similarity 56.2%; Pred. No. 0.012;  
Matches 9; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKLWGVPPYGRAXSI 16  
:|||||: :  
Db 261 NKLWVGIPYGRSFTL 276

RESULT 13  
US-09-052-778-16  
Sequence 16, Application US/09052778A  
Patent No. 6060590  
GENERAL INFORMATION:  
APPLICANT: Bryant, Peter J.  
APPLICANT: Kawamura, Kazuo  
TITLE OF INVENTION: CHITINASE RELATED PROTEINS AND METHODS  
FILE REFERENCE: 07306/015001  
CURRENT APPLICATION NUMBER: US/09/052,778A  
PRIOR FILING DATE: 1998-03-31  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 16  
LENGTH: 554  
TYPE: PRT  
ORGANISM: Manduca sexta  
US-09-052-778-16

Query Match  
Best Local Similarity 69.9%; Score 58; DB 3; Length 554;  
Best Local Similarity 56.2%; Pred. No. 0.012;  
Matches 9; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 1 DKLWGVPPYGRAXSI 16  
:|||||: :  
Db 261 NKLWVGIPYGRSFTL 276

RESULT 14  
US-09-545-814-32  
Sequence 32, Application US/09545814  
Patent No. 6416977  
GENERAL INFORMATION:  
APPLICANT: Becher, Anna M.  
TITLE OF INVENTION: FLEA CHITINASE NUCLEIC ACID MOLECULES, PROTEINS AND  
FILE REFERENCE: PC-5-C1  
CURRENT APPLICATION NUMBER: US/09/545,814  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: 60/128,833  
NUMBER OF SEQ ID NOS: 40  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 32  
LENGTH: 635

TYPE: PRT  
ORGANISM: Tagged Ctenocephalides felis  
US-09-545-814-32

Query Match 67.5%; Score 56; DB 4; Length 635;  
Best Local Similarity 83.3%; Pred. No. 0.034;  
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVGVPFYGR 12  
Db 267 KLVGVPFYGR 278

RESULT 15  
US-08-591-629-2  
Sequence 2, Application US/08591629  
Patent No. 5993808  
GENERAL INFORMATION:  
APPLICANT: MELCHERS, Leo Sjoerd  
APPLICANT: APOTHEKER-DE GROOT, Marion  
APPLICANT: BOL, John Ferdinand  
APPLICANT: CORNELISSEN, Bernardus Johannes Clemens  
APPLICANT: LINTHORST, Hubertus Josephus Maria  
APPLICANT: PONSSTEIN, Anne Silene  
APPLICANT: SEILA-BUURLAGE, Marianne Beatrice  
TITLE OF INVENTION: Plant chitinases, DNA coding therefor and  
TITLE OF INVENTION: Plants containing same  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ladae & Parry  
STREET: 26 West 61st Street  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10023-7604  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
COMPUTER: IBM PC 4.86 SX 50 Mhz  
OPERATING SYSTEM: DOS 6.20  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/591,629  
FILING DATE: 15-FEB-96  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP94/02761  
FILING DATE: 17-AUG-94  
PRIOR APPLICATION DATA: EP 93202425.0  
FILING DATE: 17-AUG-93  
ATTORNEY/AGENT INFORMATION:  
NAME: MASS, CLIFFORD J.  
REGISTRATION NUMBER: 30,086  
REFERENCE/DOCKET NUMBER: U-010627-0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 708-1800  
TELEFAX: (212) 246-8959  
TELEX: 233288  
INFORMATION FOR SEQ. ID NO. 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 371 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-591-629-2

Query Match 62.7%; Score 52; DB 2; Length 371;  
Best Local Similarity 75.0%; Pred. No. 0.1;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 KLVGVPFYGR 13  
Db 245 KLVGVPFYGR 256

Search completed: March 22, 2004, 07:03:57  
Job time: 2.37021 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:40:28 ; Search time 1.23569 Seconds

(without alignments)  
1479.047 Million cell updates/sec

Title: US-09-662-293-11

Perfect score: 109

Sequence: 1 DIPPTNIHKYLVCESYNG 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR\_78:\*\*

2: PIR1:\*\*

3: PIR3:\*\*

4: PIR4:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	56	51.4	1635	2 T14075	chitinase (EC 3.2.
2	53.5	49.1	66	2 T13181	exiclonase - Lacto
3	53	48.6	525	2 T4445	chitinase (EC 3.2.
4	47	43.1	291	2 AC1136	shikimate 5-dehydr
5	47	43.1	291	2 AE1494	shikimate 5-dehydr
6	47	43.1	330	2 S53990	DNA methyltransfer
7	47	43.1	841	1 VGBE37	glycoprotein H - h
8	46	42.2	586	2 C71695	CTP synthase (pyri
9	45	41.3	115	2 T41887	AcMNPV orf150 - Bo
10	45	41.3	342	2 T40090	hypothetical prote
11	45	41.3	504	2 A38221	chitinase (EC 3.2.
12	45	41.3	523	2 G84732	probable ubiquitin
13	45	41.3	867	2 S57795	probable deoxyribo
14	44	40.4	91	2 D97130	transcription regu
15	44	40.4	171	1 B60010	early E1A 20X prot
16	44	40.4	402	2 S42367	lag-2 protein - Ca
17	44	40.4	695	2 D86392	TIK7.9 protein - A
18	44	40.4	995	2 T05842	hypothetical prote
19	43.5	39.9	1496	2 AH0447	insecticidal toxin
20	43	39.4	102	2 F71354	probable ribosomal
21	43	39.4	304	2 T23919	hypothetical prote
22	43	39.4	495	2 S60761	alpha-2,8 polysial
23	43	39.4	502	2 D86765	hypothetical prote
24	43	39.4	683	1 A23690	protein kinase (EC
25	43	39.4	683	1 S29478	protein kinase C (
26	43	39.4	897	2 H86905	cation-transportin
27	42.5	39.0	533	2 G69735	xylan 1,4-beta-xy
28	42	38.5	76	2 AH2310	hypothetical prote
29	42	38.5	99	2 H72868	AcOrf-150 protein

30	42	38.5	116	2 UC7620	Guanylin precursor
31	42	38.5	231	2 A60468	venombin A (EC 3.4
32	42	38.5	236	1 A32121	snake venom factor
33	42	38.5	236	1 B32121	snake venom factor
34	42	38.5	254	2 T18987	hypothetical prote
35	42	38.5	317	2 S00019	L-lactate dehydrog
36	42	38.5	401	2 T24929	hypothetical prote
37	42	38.5	462	2 A72521	probable glutamine
38	42	38.5	743	2 A97021	pyruvate-formate 1
39	41.5	38.1	172	2 AB1434	spermidine N1-acet
40	41.5	38.1	183	2 H97760	hypothetical prote
41	41.5	38.1	243	2 AB3152	transcription regu
42	41.5	38.1	243	2 F98135	deor protein (AeO)
43	41.5	38.1	434	2 E97026	aspartate oxidase
44	41.5	38.1	526	2 T13484	friezzed protein h
45	41.5	36.1	703	2 S40710	hypothetical prote

## ALIGNMENTS

## RESULT 1

T14075 chitinase (EC 3.2.1.14) - yellow fever mosquito

C/Species: Aedes aegypti (yellow fever mosquito)

C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C/Accession: T14075

R/de la Vega, H.; Specht, C.A.; Liu, Y.; Robbins, P.W.

Insect Mol. Biol. 7, 233-239, 1997

A/Title: Chitinases are a multi-gene family in Aedes, Anopheles, and Drosophila.

A/Reference number: Z17872

A/Accession: T14075

A/status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1635 <DEL>

A/Cross-references: EMBL:AF026492; NID:G2564720; PID:G2564721; PIDN:AA81850.1

A/Genes: CHT2

A/Introns: 462/3; 524/3; 618/1; 951/3; 1151/2

C/Keywords: glycosidase; hydrolase; polysaccharide degradation

A/Status: preliminary

A/Accession: T13181

C/Species: Lactobacillus phage phi-gle

C/Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 13-Aug-1999

C/Accession: T13181

R/Kodaira, K.I.; Oki, M.; Kakikawa, M.; Watanabe, N.; Hirakawa, M.; Yamada, K.; Takeo,

Gene 187, 45-53, 1997

A/Title: Genome structure of the lactobacillus temperate phage phi gle: the whole genom

A/Reference number: Z17631; MUID:97225795; PMID:9073065

A/Accession: T13181

A/status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-66 <XOD>

A/Cross-references: EMBL:X98106; NID:e917136; PID:e247183; PIDN:CA66757.1

C/Genes: Xis

A/Genes: Xis

Query Match

Best Local Similarity 49.1%; Score 53.5; DB 2; Length 66;

Matches 9; Conservative 4; Mismatches 2; Indels 3; Gaps 1;

QY 1 DIPPTNI--HKYLVCE 15

Db 37 DVEPTNASHYRFLACD 54

## RESULT 3

T44445  
chikimase (BC 3.2.1.14) [imported] - African malaria mosquito  
C/Species: Anopheles gambiae (African malaria mosquito)  
C/Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jan-2000  
C/Accession: T44445  
R/Shen, Z.; Jacobs-Lorena, M.  
submitted to the EMBL Data Library, June 1997  
A/Reference number: Z22771

A/Accession: T44445  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-525 <SHR>  
A/Cross-references: EMBL:AF008575; PIDN:AAB87764.1  
A/Experimental source: adult; gut  
C/Genetics:  
A/Gene: chl-1  
C/Keywords: glycosidase; hydrolase

Query Match 48.6%; Score 53; DB 2; Length 525;  
Best Local Similarity 53.8%; Pred. No. 1.9;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 IPHPTNIHKYLC 14  
: : : : :  
Db 478 VPHPTNCAHYIC 490

## RESULT 4

shikimate 5-dehydrogenase homolog lmo0490 [imported] - Listeria monocytogenes (strain EC  
C/Species: Listeria monocytogenes  
C/Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 07-Jul-2003  
C/Accession: AC1136  
R/Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feihl, H.  
D.; Jones, L.M.; Karsc, U.  
Science 294, 849-852, 2001  
A/Authors: Krell, U.; Kuhn, M.; Kunst, F.; Kurapkac, G.; Madeno, E.; Maitournam, A.; Ma  
ok, C.; Schluter, T.; Simoes, N.; Tietzer, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,  
A./Title: Comparative genomics of Listeria species.  
A/Reference number: AB1077; MUID:21537279; PMID:11679669  
A/Accession: AC1136  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-291 <GLA>  
A/Cross-references: GB:NC 003210; PIDN:CAC98569.1; PID:gl6409866; GSPDB:GN00177  
A/Experimental source: strain EGD-e  
C/Genetics:  
A/Gene: lmo0490  
C/Superfamily: shikimate 5-dehydrogenase; shikimate dehydrogenase homology

Query Match 43.1%; Score 47; DB 2; Length 291;  
Best Local Similarity 72.7%; Pred. No. 9.1;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 IPHPTNIHKYL 12  
: : : : :  
Db 74 MPKNTIHKYL 84

## RESULT 5

shikimate 5-dehydrogenase homolog lmo0493 [imported] - Listeria innocua (strain C11p1126  
C/Species: Listeria innocua  
C/Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 07-Jul-2003  
C/Accession: AB1494  
R/Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feihl, H.

D.; Jones, L.M.; Karsc, U.  
Science 294, 849-852, 2001  
A/Authors: Krell, U.; Kuhn, M.; Kunst, F.; Kurapkac, G.; Madeno, E.; Maitournam, A.; Ma  
ok, C.; Schluter, T.; Simoes, N.; Tietzer, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,  
A./Title: Comparative genomics of Listeria species.  
A/Reference number: AB1077; MUID:21537279; PMID:11679669  
A/Accession: AB1494  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-291 <GLA>  
A/Cross-references: GB:AL592022; PIDN:CAC95725.1; PID:gl6412933; GSPDB:GN00178  
A/Experimental source: strain C11p1126  
C/Genetics:  
A/Gene: lmo0493  
C/Superfamily: shikimate 5-dehydrogenase; shikimate dehydrogenase homology

Query Match 43.1%; Score 47; DB 2; Length 291;  
Best Local Similarity 72.7%; Pred. No. 9.1;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 IPHPTNIHKYL 12  
: : : : :  
Db 74 MPKNTIHKYL 84

## RESULT 6

S53990  
DNA methyltransferase pmu1 - fission yeast (Schizosaccharomyces pombe)  
C/Species: Schizosaccharomyces pombe  
C/Date: 15-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 17-Mar-2000  
C/Accession: S53990; 139793  
R/Milkinson, C.R.M.; Bartlett, R.; Nurse, P.; Bird, A.P.  
Nucleic Acids Res. 23, 203-210, 1995  
A/Title: The fission yeast gene pmu1(+) encodes a DNA methyltransferase homologue.  
A/Reference number: S53990; MUID:9516638; PMID:7862522  
A/Accession: S53990  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-330 <WIL>  
A/Cross-references: EMBL:X82444; NID:9563910; PIDN:CA57824.1; PID:9563911  
R/McDonnell, R.C.; Rajandream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.  
submitted to the EMBL Data Library, July 1999  
A/Reference number: Z21880  
A/Accession: T39793  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-330 <WCD>  
A/Cross-references: EMBL:AL109731; PIDN:CA552029.1; GSPDB:GN00067; SPDB:SPBC19C2.02  
C/Genetics:  
A/Experimental source: strain 972h-; cosmid cl9C2  
A/Gene: SPBC19C2.02  
A/Map position: 2  
A/Inserts: 56/3  
C/Superfamily: site-specific methyltransferase (cytosine-specific) EcoRII

Query Match 43.1%; Score 47; DB 2; Length 330;  
Best Local Similarity 38.9%; Pred. No. 10;  
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 2 IPHPTNIHKYLCESYNG 19  
: : : : :  
Db 108 LPHVNNLPYILLIENVOG 125

## RESULT 7

VGB37  
glycoprotein H - human herpesvirus 3  
N/Alternate names: glycoprotein III  
C/Species: human herpesvirus 3, varicella-zoster virus  
C/Date: 30-Sep-1988 #sequence\_revision 30-Sep-1988 #text\_change 16-Jul-1999  
C/Accession: B27341  
R/Davison, A.J.; Scott, J.E.  
J. Gen. Virol. 67, 1759-1816, 1986

A>Title: The complete DNA sequence of varicella-zoster virus.  
A/Reference number: A27345; MUID:86306657; PMID:3018124

A/Accession: B27341

A/Molecule type: DNA

A/Residues: 1-841 <DAV>

A/Cross-references: EMBL:X04370; NID:955989; PIDN:CAA27920.1; PID:960026

C/Genetics:

A/Gene: 37

C/Superfamily: herpesvirus glycoprotein H

C/Keywords: glycoprotein

F,18,45,83,217,317,499,522,560,760,783/Binding site: carbohydrate (asn) (covalent) #stat

Query Match 43.1%; Score 47; DB 1; Length 841;  
Best Local Similarity 50.0%; Pred. No. 27;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 IPHPTNKHVLCESV 17  
DB 715 LPHPTNKHVLCESV 730

RESULT 8

CT1695

CTP synthase (pyrg) RP378 - Rickettsia prowazekii

C/Species: Rickettsia prowazekii

C/Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000

C/Accession: C71695

R/Anderson, S.G.E.; Zomorodipour, A.; Anderson, J.O.; Sicheit-Ponten, T.; Almark, U

Nature 396, 133-140, 1998

A/Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.

A/Reference number: A71630; MUID:99039499; PMID:9823893

A/Accession: C71695

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-586 <AND>

A/Cross-references: GB:A235271; GB:A235269; NID:93868717; PIDN:CAA14837.1; PID:9386093

A/Experimental source: strain Madrid E

C/Genetics:

A/Gene: pyrg; RP378

C/Superfamily: CTP synthase

Query Match 42.2%; Score 46; DB 2; Length 586;  
Best Local Similarity 50.0%; Pred. No. 27;  
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 IPHPTNKHVLCESV 19  
DB 115 IPHPTNKHVLCESV 132

RESULT 9

T41887

ACMNV orf150 - Bombyx mori nuclear polyhedrosis virus (isolate T3)

C/Species: Bombyx mori nuclear polyhedrosis virus, BMSNPV

A/Variety: isolate T3

C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 11-May-2000

C/Accession: T41887

R/Gomi, S.; Majima, K.; Maeda, S.

J. Gen. Virol. 80, 1323-1337, 1999

A/Title: Sequence analysis of the genome of Bombyx mori nucleopolyhedrovirus.

A/Reference number: Z22020; MUID:95281911; PMID:10355780

A/Accession: T41887

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-115 <KAM>

A/Cross-references: EMBL:U33180; PIDN:AA63816.1

A/Experimental source: isolate T3

C/Genetics:

A/Note: Orf\_126

QY 3 PHPTNKHVLCESV 18  
DB 60 PHPTNKHVLCESV 75

RESULT 10

T40090

hypothetical protein SPBC29A3.19 - fission yeast (Schizosaccharomyces pombe)

C/Species: Schizosaccharomyces pombe

C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 20-Oct-2000

C/Accession: T40090; T39752

R/Lyne, M.; Rajandream, K.A.; Barrell, B.G.; Volckaert, G.

submitted to the EMBL Data Library, March 1998

A/Reference number: Z21904

A/Accession: T40090

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-342 <LYN>

A/Cross-references: EMBL:AL022299; PIDN:CAA18396.1; GSPDB:GN00067; SPDB:SPBC29A3.19

A/Experimental source: strain 972h-; cosmid c29A3

R/Mood, V.; Barrell, B.G.; Rajandream, M.A.; Xiang, Z.; Aves, S.

submitted to the EMBL Data Library, January 1999

A/Accession: T39752

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 198-342 <MOO>

A/Cross-references: EMBL:AL035077; PIDN:CAA22661.1; GSPDB:GN00066; SPDB:SPBC18B5.01

A/Experimental source: strain 972h-; cosmid c18B5

C/Genetics: <LYN>

A/Gene: SPDB:SPBC29A3.19

A/Map position: 2

C/Genetics: <MOO>

A/Gene: SPDB:SPBC18B5.01

A/Map position: 1

Query Match 41.3%; Score 45; DB 2; Length 342;  
Best Local Similarity 58.3%; Pred. No. 22;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 PHPTNKHVLC 14  
DB 130 PHPTNKHVLC 141

RESULT 11

A38221

chitinase (EC 3.2.1.14) Mf1 - nematode (Brugia malayi)

C/Species: Brugia malayi

C/Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 21-Jul-2000

C/Accession: A38221

R/Fuhrman, J.A.; Lane, W.S.; Smith, R.F.; Plessens, W.F.; Perlter, F.B.

Proc. Natl. Acad. Sci. U.S.A. 89, 1548-1552, 1992

A/Title: Transmission-blocking antibodies recognize microfilarial chitinase in brugian

A/Reference number: A38221; MUID:9219220; PMID:1542646

A/Accession: A38221

A/Status: preliminary; not compared with conceptual translation

A/Molecule type: nucleic acid; protein

A/Residues: 1-504 <FUH>

A/Cross-references: GB:M73689; NID:9156063; PIDN:AA27854.1; PID:9156064

A/Note: sequence extracted from NCBI backbone (NCBI:85345)

C/Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 41.3%; Score 45; DB 2; Length 504;  
Best Local Similarity 50.0%; Pred. No. 33;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 PHPTNKHVLC 14  
DB 459 PHPTNKHVLC 470

## RESULT 12

G84732  
Probable ubiquitin activating enzyme [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: G84732  
R/Lin: X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.T.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, W.; VanAken, S.E.; Umayam, L.; Tallon, L.  
euser, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: G84732  
A>Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-523 <STO>  
A:Cross-references: GB:AE002093; NID:G3831455; PIDN:AA069937.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g32410  
A:Map position: 2

Query Match 41.3%; Score 45; DB 2; Length 523;  
Best Local Similarity 47.1%; Pred. No. 35;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 PPHPTNKHVLCESVNG 19  
DB 393 PSTTEHKLHVLADENYSG 409

## RESULT 13

SS7795  
Probable deoxyribodipyrimidine photo-lyase (EC 4.1.99.3) - Chlamydomonas reinhardtii  
N:Alternate names: DNA photolyase homolog; probable blue light photoreceptor  
C:Species: Chlamydomonas reinhardtii  
C:Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 21-Jul-2000  
C:Accession: S57795; S66368  
R/Small: G.D.; Min, B.; Lefebvre, P.A.  
Plant Mol. Biol. 28, 443-454, 1995  
A:Title: Characterization of a Chlamydomonas reinhardtii gene encoding a protein of the  
A:Reference number: S57795; MUID:95359403; PMID:7632915  
A:Accession: S57795  
A:Molecule type: DNA  
A:Residues: 1-867 <SMA>  
A:Cross-references: EMBL:L07561; NID:G945420; PIDN:AA037438.1; PID:G945421  
A:Accession: S66368  
A:Molecule type: mRNA  
A:Residues: 1-867 <SNM>  
A:Cross-references: EMBL:L07561; NID:G945420; PIDN:AA037438.1; PID:G945421  
C:Genetics:  
A:Gene: CPH1  
A:Introns: 34/3, 159/2, 210/3, 265/3, 329/3, 406/1, 837/3  
C:Keywords: carbon-carbon lyase; photoreceptor

Query Match 41.3%; Score 45; DB 2; Length 867;  
Best Local Similarity 88.9%; Pred. No. 58;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 TNNHXYLVC 14  
DB 856 TNNHXYLVC 864

## RESULT 14

D97130  
transcription regulator (phage-related) (Xre family) [imported] - Clostridium acetobutylicum  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C:Accession: D97130  
R/Noilling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: D97130  
A>Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-91 <KUR>  
A:Cross-references: GB:AE001437; PIDN:AA079831.1; PID:G15024845; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC1867

Query Match 40.4%; Score 44; DB 2; Length 91;  
Best Local Similarity 41.2%; Pred. No. 8.2;  
Matches 7; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 2 IHPPTNKHVLCESVN 18  
DB 70 IHPPTNKHVLCESVN 84

## RESULT 15

B60010  
early E1A 20K protein - canine adenovirus 2  
C:Species: Mastadenovirus can2 (canine adenovirus 2)  
A:Note: host Canis lupis familiaris (dog)  
C:Date: 17-Feb-1994 #sequence\_revision 17-Feb-1994 #text\_change 07-May-1999  
C:Accession: B60010  
R/Spidey, N.; McClure, R.S.; Cavanagh, H.M.A.  
Virus Res. 14, 241-256, 1989  
A:Title: Identification and nucleotide sequence of the early region 1 from canine adenov.  
A:Reference number: B60010; MUID:90163565; PMID:2623943  
A:Accession: B60010  
A:Molecule type: DNA  
A:Residues: 1-171 <SPI>  
C:Superfamily: adenovirus early E1A protein  
C:Keywords: early protein; transcription regulation

Query Match 40.4%; Score 44; DB 1; Length 171;  
Best Local Similarity 30.0%; Pred. No. 16;  
Matches 9; Conservative 4; Mismatches 5; Indels 12; Gaps 1;

QY 2 IHPPTNKHVLCESVN 19  
DB 6 VPAPRTHDYVLELEWHPDCLDCEYNG 35

Search completed: March 22, 2004, 07:01:27  
Job time : 2.23569 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:31:13 ; Search time 0.728738 Seconds  
(without alignment)  
1357.597 Million cell updates/sec

Title: US-09-662-293-11  
Perfect score: 109  
Sequence: 1 DIPHPNTHKYLVCSSVNG 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 5207015 residues  
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	43.1	291	1 AROE_LISIN	Q926G7 listeria in
2	47	43.1	291	1 AROE_LISIN	Q926G7 listeria in
3	47	43.1	330	1 PMTI_SCHPO	P40939 schizosacch
4	47	43.1	841	1 VGLH_VZVD	P09360 varicella-z
5	46	42.2	153	1 PBLI_ANOGA	O76317 anopheles g
6	46	42.2	586	1 PYRG_RICPR	O92611 rickettsia
7	45	41.3	342	1 YBSI_SCHPO	O59681 schizosacch
8	45	41.3	458	1 CH13_DROME	O9W630 drosophila
9	45	41.3	504	1 CH1T_BRUMA	P29030 brugia mala
10	44	40.4	103	1 R310_FUSNN	O8r1f4 fusobacteri
11	44	40.4	171	1 E1A_ADEEC2	P35881 canine aden
12	44	40.4	402	1 LAG2_CAEEL	P45342 caenorhabdi
13	43	39.4	102	1 R310_TREPA	O83345 treponema p
14	43	39.4	258	1 VSP3_TRIGA	O13063 trimersoma p
15	43	39.4	434	1 BRAC_HEMPU	O25113 hemelcentro
16	43	39.4	683	1 KPCL_MOUSE	P23398 mus muscultu
17	43	39.4	683	1 KPCL_MOUSE	O64617 rattus norv
18	42.5	39.0	935	1 AD22_XENLA	O41296 xenopus lae
19	42	38.5	99	1 Y150_NPVAC	P41107 autocographa
20	42	38.5	231	1 VSP1_AKGC	P09672 agkistrodon
21	42	38.5	236	1 VSP1_AKGC	P18664 dabolia russ
22	42	38.5	236	1 VSP1_AKGC	P18665 dabolia russ
23	42	38.5	252	1 TPIS_STRAS	O8e644 streptococc
24	42	38.5	252	1 TPIS_STRAS	O8e644 streptococc
25	42	38.5	252	1 TPIS_STRAS	O8e644 streptococc
26	42	38.5	260	1 VSP2_VIPLE	O9p140 vicia lebe
27	42	38.5	317	1 LDH_BACCA	P10555 bacillus ca
28	42	38.5	518	1 ASB3_HUMAN	O9y575 homo sapien
29	42	38.5	1934	1 PNO_CRYPV	O968x7 cryptospori
30	41.5	38.1	434	1 NADB_CLOAB	O97K95 clostridium
31	41.5	38.1	581	1 FRZ3_DROME	O77438 drosophila
32	41.5	38.1	703	1 YK16_CAEEL	P34317 caenorhabdi
33	41	37.6	103	1 RS10_BORBU	P94266 borrelia bu

34	41	37.6	202	1 ADEN_ADEB7	P19151 bovine aden
35	41	37.6	251	1 TPIS_STRP8	Q6p143 streptococc
36	41	37.6	251	1 TPIS_STRPY	P82478 streptococc
37	41	37.6	252	1 TPIS_STRPN	Q97p11 streptococc
38	41	37.6	257	1 VSP2_BOTUA	O13069 bothrops ja
39	41	37.6	258	1 VSP3_BOTUA	O9p148 bothrops ja
40	41	37.6	260	1 VSP6_TRIMU	Q9d683 trimersoma
41	41	37.6	290	1 KDSA_ARATH	Q9a97 a.2-denydro
42	41	37.6	290	1 KDSA_PEA	O50044 pisum sativ
43	41	37.6	576	1 CBPS_YEAST	P27614 saccharomyc
44	41	37.6	926	1 RPM1_ARATH	O39214 arabidopsis
45	40.5	37.2	232	1 VSPA_BOTUA	P81661 bothrops ja

## ALIGNMENTS

RESULT 1  
AC Q926G7; ID AROE\_LISIN STANDARD; PRT; 291 AA.  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Shikimate 5-dehydrogenase (EC 1.1.1.25).  
GN AROE OR LIN0493.  
OS Listeria innocua.  
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
OX NCBI\_TaxId=1642;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CLIP 11262 / Serovar ea;  
RX MEDLINE=21537279; PubMed=11679669;  
RA Blaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
RA Gao F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
RA Charbit A., Chetoui F., Couve E., de Darivar A., Dehoux P.,  
RA Domann E., Dominguez-Bernal G., Duchaud E., Durrant L., Dussurget O.,  
RA Entian K.-D., Feil H., Garcia-del Portillo F., Garrido P.,  
RA Gautier L., Goebel W., Gomez-Iopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-W., Kaerst U., Kreft J., Kuhn M., Kunst F., Kuraprat G.,  
RA Madueno E., Maitournam A., Mala Vicente J., Ng E., Nedjari H.,  
RA Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
RA Remmel B., Rose M., Schluter T., Simoes N., Tixeront A.,  
RA Vazquez-Boland J.-A., Voss H., Weiland J., Coessart P.,  
RA "Comparative genomics of Listeria species."  
RA Science 294:849-852(2001).  
CC -1- CATALYTIC ACTIVITY: Shikimate + NADP(+) = 5-dehydroshikimate + NADPH.  
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;  
CC -1- fourth step.  
CC -1- SHIKIMATE: belongs to the shikimate dehydrogenase family.  
CC -1- SHIKIMATE: belongs to the shikimate dehydrogenase family.  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL, AL596165; CAC95725.1; -  
CC PIR, A1494; A1494.  
CC L1494; L1494.  
CC HAMAP, MF\_00222; -1.  
CC InterPro, IPR006151; Shikimate DH.  
CC Pfam, PF01488; Shikimate DH; 1\_  
CC Aromatic amino acid biosynthesis; Oxidoreductase; NADP;  
CC Complete proteome.  
SQ SEQUENCE 291 AA; 32192 MW; 119148B1BDA3F244 CRC64;  
Query Match 43.1%; Score 47; DB 1; Length 291;  
Best Local Similarity 72.7%; Pred. No. 2.9;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 IPHPTNTHKYL 12  
DB 74 MPNKTNTHKYL 84

## RESULT 2

AROELISMO STANDARD; PRT: 291 AA.

AC Q8Y9N5;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Shikimate 5-dehydrogenase (EC 1.1.1.25).  
GN AROE OR LMO0490.  
OS Listeria monocytogenes.  
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
OX NCBI\_TaxID=1639;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=EBD-e / Serovar 1/2a;  
RX MEDLINE=21537279; PubMed=11679669;  
RA Glaeser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
RA Domnan E., Dominguez-Bernal G., Duchaud E., Durant L., Duseurget O.,  
RA Entian K.-D., Feihl H., Garcia-del Portillo F., Garrido P.,  
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,  
RA Madueno E., Maizumi A., Mera Vicente J., Ng E., Nedjari H.,  
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Porcell R.,  
RA Remmel B., Rose M., Schlueter T., Simoes N., Tietze A.,  
RA Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.;  
RT "Comparative genomics of *Listeria* species.";  
RL Science 294:849-852(2001).

CC -1- CATALYTIC ACTIVITY: Shikimate + NADP(+) = 5-dehydroshikimate +  
CC NADPH;  
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;  
CC fourth step.  
CC -1- SIMILARITY: Belongs to the shikimate dehydrogenase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; AL591975; CAC98569.1; -;  
DR PIR; AC1136; AC1136.  
DR Listlist; LMO00490; -;  
DR HAMAP; MF\_00222; -;  
DR InterPro; IPR006151; Shikimate\_DH.  
DR Pfam; PF01488; Shikimate\_DH; 1.  
DR Aromatic amino acid biosynthesis; Oxidoreductase; NADP;  
KW Complete proteome.  
SQ SEQUENCE 291 AA; 32161 MW; BDBDFP2732CD7D5D CRC64;

Query Match 43.1%; Score 47; DB 1; Length 291;  
Best Local Similarity 72.7%; Pred. No. 2.9;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 IPHPTNTHKYL 12  
DB 74 MPNKTNTHKYL 84

## RESULT 3

PMT1\_SCHPO STANDARD; PRT: 330 AA.  
AC P40999;  
DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE DNA methyltransferase homolog pmc1 (SPIM.Spo1) (M.Spm1).  
GN PMT1 OR SPC19C2.02

OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.

RC STRAIN=972;  
RX MEDLINE=9516638; PubMed=7862522;  
RA Wilkinson C.R.M., Bartlett R., Nurse P., Bird A.P.;  
RT "The fission yeast gene *pmc1+* encodes a DNA methyltransferase  
RT homologue.";  
RL Nucleic Acids Res. 23:203-210(1995).  
RN [2]  
RP SEQUENCE FROM N.A.

RC STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgourou J., Peat N., Hayles S., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.U., Hunt S., Jagsels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor J., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodard J., Wolckstein G., Aert R., Robben J., Grymoprez B.,  
RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moesli D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann M., Wedler H., Wandut R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaune V., Motier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hux S.M.,  
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garçon A., Thode G.,  
RA Dugas R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cernutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;  
RT "The genome sequence of *Schizosaccharomyces pombe*.";  
RL Nature 415:871-880(2002).  
RN [3]  
RP CHARACTERIZATION.

RX MEDLINE=96194447; PubMed=8636983;  
RA Pinarbasi E., Elliott J., Horny D.P.;  
RT "Activation of a yeast pseudo DNA methyltransferase by deletion of a  
RT single amino acid.";  
RL J. Mol. Biol. 257:804-813(1996).

CC -1- FUNCTION: Does not have a cytosine-5 methyltransferase activity  
CC found at the insertion of a Ser residue between the Pro-Cys motif  
CC it becomes catalytically active and recognizes and methylates the  
CC sequence CC[AT]GG.  
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -1- SIMILARITY: Belongs to the C5-methyltransferase family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; X82444; CAA57824.1; -;  
DR EMBL; AL109731; CAB52023.1; -;  
DR PIR; S53990; S53990.



DR HSP; 014717; 1G55.  
 DR REBASE; 2888; M.Spomt.  
 DR GenDB; SPombe; SPC19C2.02; "-  
 DR InterPro; IP001525; C5\_DNA\_mech.  
 DR Pfam; PF00145; DNA\_methylase; 1.  
 DR PRINTS; PR00105; C5METTRFASB.  
 DR PROSITE; PS00094; C5\_MTASE\_1; FALSE\_NEG.  
 DR PROSITE; PS00095; C5\_MTASE\_2; 1.  
 KW DNA-binding; Nuclear protein.  
 FT SITE 81 81 ANCESTRAL ACTIVE SITE.  
 SQ SEQUENCE 330 AA; 37976 MW; 50A712FAVCF38A1 CRC64;

Query Match 43.1%; Score 47; DB 1; Length 330;  
 Best Local Similarity 38.9%; Pred. No. 3.3;  
 Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 2 IPHPTNHHKYLVCESVNG 19  
 DB 108 LPHVNNLPVYLLENVOG 125

RESULT 4  
 VGLH VZVD STANDARD; PRT; 841 AA.  
 AC P09260;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Probable glycoprotein H precursor (Glycoprotein III) (GP111).  
 GN 37.  
 OS Varicella-zoster virus (strain Dumas) (VZV).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Varicellovirinae;  
 NCBI\_TaxID=10338;  
 [1]  
 RX MEDLINE=86306657; PubMed=3018124;  
 RA Davison A.U.; Scott J.E.;  
 RL "The complete DNA sequence of varicella-zoster virus,"  
 J. Gen. Virol. 67:1759-1816(1986).  
 CC -1- SIMILARITY: Belongs to the herpesviruses glycoprotein H family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X04370; CAA27920.1; -  
 DR PIR; B27341; VGBE37.  
 KW Glycoprotein; Transmembrane; Signal.  
 FT SIGNAL 1 17  
 FT CHAIN 1 841  
 FT CARBOHYD 18 18 PROBABLE GLYCOPROTEIN H.  
 FT CARBOHYD 45 45 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 217 217 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 317 317 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 760 760 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 783 783 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 841 AA; 93651 MW; 82B247F63CA51948 CRC64;

Query Match 43.1%; Score 47; DB 1; Length 841;  
 Best Local Similarity 50.0%; Pred. No. 9.2;  
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 IPHPTNHHKYLVCESV 17  
 DB 715 LPHPTNHHKYLVCESV 730

RESULT 5  
 ID PBL\_ANOGA STANDARD; PRT; 153 AA.  
 AC 076217;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Peritrophin-1 precursor.  
 GN Ape1.  
 OS Anopheles gambiae (African malaria mosquito).  
 CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 CC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.  
 NCBI\_TaxID=7165;  
 [1]  
 RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.  
 RC STRAIN=G3; TISSUE=Midgut;  
 RX MEDLINE=9831635; PubMed=9651363;  
 RA Shen Z.; Jacobs-Lorena M.;  
 RT "A type I peritrophic matrix protein from the malaria vector Anopheles  
 RT gambiae binds to chitin. Cloning, expression, and characterization.";  
 RL J. Biol. Chem. 273:17665-17670(1998).  
 CC -1- FUNCTION: Binds chitin but not cellulose. May be involved in the  
 CC spatial organization of PM.  
 CC -1- TISSUE SPECIFICITY: Adult peritrophic membrane.  
 CC -1- DEVELOPMENTAL STAGE: Expressed in adult but not larval guts,  
 CC whole pupae or whole bodies minus gut.  
 CC -1- PM: Glycosylated.  
 CC -1- SIMILARITY: Contains 2 chitin-binding type-2 domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF030431; AAC39127.1; -  
 DR InterPro; IPR002557; Chitin\_bind\_Pera.  
 DR Pfam; PF01607; CBM\_14; 2.  
 DR SMART; SM00494; CHIT\_BIND\_2.  
 DR PROSITE; PS50940; CHIT\_BIND\_1; 2.  
 KW Chitin-binding; Glycoprotein; Repeat; Signal.  
 FT SIGNAL 1 17  
 FT CHAIN 18 153 PERITROPHIN-1.  
 FT DOMAIN 18 79 CHITIN-BINDING TYPE-2.1.  
 FT DOMAIN 92 153 CHITIN-BINDING TYPE-2.2.  
 FT CARBOHYD 63 63 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 153 AA; 16819 MW; 221BDA0310476338 CRC64;

Query Match 42.8%; Score 46; DB 1; Length 153;  
 Best Local Similarity 50.0%; Pred. No. 2.1;  
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 IPHPTNHHKYLVCESV 15  
 DB 107 IPHPTNHHKYLVCESV 120

RESULT 6  
 ID PYRG\_RICPR STANDARD; PRT; 586 AA.  
 AC Q9ZDF1;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase) (CTP synthetase).  
 GN PYRG OR RP378.  
 OS Rickettsia prowazekii.  
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
 CC Rickettsiaceae; Rickettsia; Rickettsia;  
 NCBI\_TaxID=782;

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid B;
RX MEDLINE=99039439; PubMed=9823893;
RA Andersson S.G.E., Zomorodi-pour A., Andersson J.O.,
RA Scheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
RN
RP DOMAIN REPI.
RX MEDLINE=20485642; PubMed=11030655;
RA Ocata H., Audic S., Barbe V., Artiguenave F., Fournier P.-E.,
RA Raulot D., Claverie J.-M.;
RT "Selfish DNA in protein-coding genes of Rickettsia.";
RL Science 290:347-350(2000).
CC -1- FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with
CC either L-glutamine or ammonia as the source of nitrogen (By
CC similarity).
CC -1- CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.
CC -1- ENZYME REGULATION: Allosterically activated by GTP, when glutamine
CC is the substrate. Inhibited by CTP (By similarity).
CC -1- PATHWAY: Pyrimidine biosynthesis; conversion of UTP to CTP; third
CC (last) step.
CC -1- SUBUNIT: Homotrimer (By similarity).
CC -1- SIMILARITY: Belongs to the CTP synthase family.
CC -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
CC -1- SIMILARITY: Contains 1 REPI insert domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AJ235271; CAA14837.1; -.
CC PIR: C71695; C71695.
CC DR HAMAP: MF_01227; acypical; 1.
CC DR InterPro: IPR000991; GATase.1.
CC DR InterPro: IPR004468; Pyrg_synth.
CC DR InterPro: IPR005728; Ricketc_RP2.
CC DR Pfam: PF00117; GATase; 1.
CC DR TIGRFAMs: TIGR00337; Pyrg; 1.
CC DR TIGRFAMs: TIGR01045; RPB; 1.
CC DR PROSITE: PS00442; GATASE_TYPE_1; 1.
CC KM Pyrimidine biosynthesis; ligase; Glutamine amidotransferase;
CC Complete proteome.
CC FT DOMAIN 1 299 AMINATOR DOMAIN.
CC FT DOMAIN 300 586 GLUTAMINE AMIDOTRANSFERASE.
CC FT DOMAIN 429 473 REPI INSERT.
CC FT ACT_SITE 379 379 GATASE (BY SIMILARITY).
CC FT ACT_SITE 555 555 GATASE (BY SIMILARITY).
CC FT ACT_SITE 557 557 GATASE (BY SIMILARITY).
CC SQ SEQUENCE 586 AA; 66173 MW; FA14C0879F457A0A CRC64;
QY
QY Query Match 42.2%; Score 46; DB 1; Length 586;
QY Best Local Similarity 50.0%; Pred. No. 9;
QY Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
Db 115 IPHVTNTHKLVYCESVNG 132

```

```

DT 28-FEB-2003 (Rel. 41, last annotation update)
DE Hypothetical protein C18E5.01 in chromosome II.
GN SPBC18E5.01 OR SPBC29A3.19.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OC NCBI_TaxID=4896;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Brooks K., Beat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Sgouros K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hildalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.D., Hunt S., Jagers K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymptre B.,
RA Welfens I., Vansirels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moesli D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler S., Gloux S., Lelaue V., Mottier S.,
RA Goffeau A., Cadieu E., Dreano S., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Galbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado I., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Folsburg S.L.,
RA Cernutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakowski G.V., Ussery D., Barrett B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -1- SIMILARITY: Belongs to the cycloisomerase 2 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AL022299; CAA18396.1; -.
CC DR EMBL: AL035077; CAA22661.1; -.
CC DR PIR: T40090; T40090.
CC DR GeneDB: SPombe; SPBC18E5.01; -.
CC FT Hypothetical protein.
CC FT DOMAIN 153 168 POLY-VAL.
CC SQ SEQUENCE 342 AA; 38220 MW; AE39A39B053F704F CRC64;
QY
QY Query Match 41.3%; Score 45; DB 1; Length 342;
QY Best Local Similarity 58.3%; Pred. No. 7.3;
QY Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
Db 130 PHQVTAHKLVC 141

```

```

RESULT 7
YBS1_SCHPO STANDARD; PRT; 342 AA.
AC OS9681; OSUSXS;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, last sequence update)

```

```

RESULT 8
CH13_DROME STANDARD; PRT; 458 AA.
AC Q9W5U2; O17422;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, last sequence update)
DT 15-MAR-2004 (Rel. 43, last annotation update)
DE Probable chitinase 3 (EC 3.2.1.14).

```

GN CHT3 OR CG18140.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RX MEDLINE=2246071; PubMed=12537574;  
RA Hoskins R.A., Smith C.D., Carlson J.W., Carvalho A.B., Halpern A.,  
RA Kaminker J.S., Kennedy C., Mungall C.W., Sullivan B.A., Sutton G.G.,  
RA Yasuda R.J.C., Wakimoto B.T., Myers E.W., Celisner S.E., Rubin G.M.,  
RA Karpis G.H.;  
RT "Heterochromatic sequences in a Drosophila whole-genome shotgun  
RT assembly";  
RL Genome Biol. 3:RESEARCH0085.1-RESEARCH0085.16(2002).  
RN [2]  
RP SEQUENCE OF 182-294 FROM N.A.  
RC STRAIN=Canton-S;  
RX MEDLINE=98324849; PubMed=9662472;  
RA de la Vega H., Specht C.A., Liu Y., Robbins P.W.;  
RT "Chitinases are a multi-gene family in Aedes, Anopheles and  
RT Drosophila";  
RL Insect Mol. Biol. 7:233-239(1998).  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-  
CC acetyl-D-glucosamine polymers of chitin.  
CC -1- SIMILARITY: Belongs to chitinase class II (family 18 of glycosyl  
CC hydrolases).  
CC -1- SIMILARITY: Contains 2 chitin-binding type-2 domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF026502; AAB81860.1; -  
DR FlyBase; FBgn0022701; Cht3.  
DR InterPro; IPR002557; Chitin bind PerA.  
DR InterPro; IPR001223; Glyco\_hydro\_18.  
DR InterPro; IPR001579; Glyco\_hydro\_18AS.  
DR Pfam; PF01607; CBM\_14; 2.  
DR Pfam; PF00704; Glyco\_hydro\_18; 1.  
DR ProDom; PD000471; Glyco\_hydro\_18; 1.  
DR SMART; SM00494; Chtd2; 2.  
DR SMART; SM00636; Glyco\_18; 1.  
DR PROSITE; PS50940; CHIT\_BIND\_II; 2.  
DR PROSITE; PS01095; CHITINASE\_18; FALSE NEG.  
KW Hydrolyase; Glycosidase; Chitin degradation; Chitin-binding;  
KW Multigene family; Repeat.  
FT DOMAIN 5 58  
FT ACT\_SITE 74 128 CHITIN-BINDING TYPE-2 1.  
FT ACT\_SITE 295 295 CHITIN-BINDING TYPE-2 2.  
SQ SEQUENCE 458 AA; 52330 MW; 4A063190B7E96248 CRC64;  
Query Match 41.3%; Score 45; DB 1; Length 458;  
Best Local Similarity 53.8%; Pred. No. 10;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DB Endochitinase precursor (Ec 3.2.1.14) (MFI antigen).  
OS Brugia malayi (Filarial nematode worm).  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;  
OC Onchocercidae; Brugia.  
OX NCBI\_TaxID=6279;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC MEDLINE=92179220; PubMed=1542646;  
RA Fuhrman J.A., Lane W.S., Smith R.F., Plessens W.F., Perler F.B.;  
RT "Transmission-blocking antibodies recognize microfilarial chitinase  
RT in human lymphatic filariasis";  
RT Proc. Natl. Acad. Sci. U.S.A. 89:1548-1552(1992).  
CC -1- FUNCTION: Microfilarial chitinase, which may function to degrade  
CC chitin-containing structures in the micro-filaria or in its  
CC mosquito vector during parasite development and transmission.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-  
CC acetyl-D-glucosamine polymers of chitin.  
CC -1- DEVELOPMENTAL STAGE: The appearance of the MFI antigen correspond  
CC with the onset of the parasite's ability to infect the mosquito.  
CC -1- PTM: O-glycosylated.  
CC -1- MISCELLANEOUS: Known to bind calcium.  
CC -1- SIMILARITY: Belongs to chitinase class II (family 18 of glycosyl  
CC hydrolases).  
CC -1- SIMILARITY: Contains 1 chitin-binding type-2 domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M73689; AAA27854.1; -  
DR PIR; A38221; A38221.  
DR InterPro; IPR002557; Chitin bind PerA.  
DR InterPro; IPR001223; Glyco\_hydro\_18.  
DR InterPro; IPR001579; Glyco\_hydro\_18AS.  
DR Pfam; PF01607; CBM\_14; 1.  
DR Pfam; PF00704; Glyco\_hydro\_18; 1.  
DR ProDom; PD000471; Glyco\_hydro\_18; 1.  
DR SMART; SM00494; Chtd2; 1.  
DR SMART; SM00636; Glyco\_18; 1.  
DR PROSITE; PS50940; CHIT\_BIND\_II; 1.  
DR PROSITE; PS01095; CHITINASE\_18; 1.  
KW Hydrolyase; Glycosidase; Chitin degradation; Chitin-binding; Signal;  
KW Antigen; Repeat; Glycoprotein; Calcium-binding.  
FT SIGNAL 1 22  
FT CHAIN 23 504 ENDOCHITINASE.  
FT DOMAIN 23 400 CATALYTIC.  
FT DOMAIN 401 448 SER/THR-RICH (LINKER).  
FT DOMAIN 407 448 3 X 14 AA APPROXIMATE TANDEM REPEATS.  
FT DOMAIN 448 504 CHITIN-BINDING TYPE-2.  
FT ACT\_SITE 148 148 PROTON DONOR (BY SIMILARITY).  
SQ SEQUENCE 504 AA; 55971 MW; A78B7BFB83709B CRC64;  
Query Match 41.3%; Score 45; DB 1; Length 504;  
Best Local Similarity 50.0%; Pred. No. 11;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

RESULT 9  
CHIT\_BRUMA STANDARD; PRT; 504 AA.  
ID CHIT\_BRUMA  
AC P29030;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)

RESULT 10  
RS10\_FUSNN STANDARD; PRT; 103 AA.  
ID RS10\_FUSNN  
AC Q8R1F4;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE 30S ribosomal protein S10.  
 GN RPSJ OR FN1646.  
 OS Fusobacterium nucleatum (subsp. nucleatum).  
 CC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;  
 CC Fusobacterium.  
 CC NCBI\_TaxID=76856;  
 CC [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=ATCC 25586;  
 RX MEDLINE=2186394; PubMed=11889109;  
 RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,  
 RA Bhattacharya A., Barman A., Gardner W., Grechkin G., Zhu L.,  
 RA Vasileva O., Chu L., Kogan Y., Chaga O., Goldsman E., Bernal A.,  
 RA Larsen N., D'Souza M., Malinas T., Pusch G., Haselkorn R.,  
 RA Forsberg M., Kyridis N., Overbeek R.,  
 RT "Genome sequence and analysis of the oral bacterium Fusobacterium  
 nucleatum strain ATCC 25586."  
 RL J. Bacteriol. 184:2005-2018(2002).  
 CC -1- FUNCTION: Involved in the binding of tRNA to the ribosomes (by  
 CC similarity).  
 CC -1- SIMILARITY: Belongs to the S10P family of ribosomal proteins.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AE010471; ALP3761.1; -.  
 DR HAMAP: MF\_00508; -1.  
 DR InterPro: IPR001848; Ribosomal\_S10.  
 DR InterPro: IPR005731; Ribosomal\_S10\_B.  
 DR Pfam: PF00338; Ribosomal\_S10; 1.  
 DR PRINTS: PR00971; RIBOSOMALS10.  
 DR ProDom: PD001272; Ribosomal\_S10; 1.  
 DR TIGRfam: TIGR01049; rpsJ; 1.  
 DR PROSITE: PS00361; RIBOSOMAL\_S10; 1.  
 KW Ribosomal protein; Complete proteome.  
 KM RiboSOMAL\_S10; 1.  
 SQ SEQUENCE 103 AA; 11541 MW; EFSAA895BBF63DCA CRC64;  
 Query March 40.4%; Score 44; DB 1; Length 103;  
 Best Local Similarity 52.9%; Pred. No. 2.9;  
 Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Query Match 40.4%; Score 44; DB 1; Length 171;  
 Best Local Similarity 30.0%; Pred. No. 5;  
 Matches 9; Conservative 4; Mismatches 5; Indels 12; Gaps 1;  
 QY 2 IPHPTNIRKYLIV-----CESYNG 19  
 DB 6 VPAPRNHDIVLELLEEMWPPDCEIYPNG 35  
 RESULT 12  
 LAG2\_CAEEL STANDARD; PRT; 402 AA.  
 AC P43442;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE LAG-2 protein precursor  
 GN LAG-2 OR LET-461 OR Y73C8B.4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 CC NCBI\_TaxID=6239;  
 CC [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=Bristol N2.  
 RX MEDLINE=94187845; PubMed=8139658;  
 RA Tax F.E., Yeagers J.J., Thomas J.H.;  
 RT "Sequence of C. elegans lag-2 reveals a cell-signaling domain shared  
 RT with Delta and Serrate of Drosophila."  
 RL Nature 368:150-154(1994).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=Bristol N2;  
 RA Demis G., Wohldmann P., Courtney L., Drone K.;  
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RN REVISIONS.  
 RA Waterston R.;  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: PUTATIVE INTERCELLULAR SIGNAL FOR LIN-12 AND GLP-1  
 CC RECEPTORS.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- DOMAIN: The DSL domain is required for binding to the Notch  
 CC receptor.  
 CC -1- SIMILARITY: Contains 2 EGF-like domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X77495; CA54629.1; -.  
 DR EMBL: AC024205; AAF36047.1; -.  
 DR PIR: S42367; S42367.  
 DR WormPep: Y73C8B.4; CE22970.  
 DR InterPro: IPR001774; DSL.  
 DR InterPro: IPR006209; EGF-like.  
 DR InterPro: IPR006210; IEGF.  
 DR Pfam: PF01414; DSL; 1.  
 DR Pfam: PF00008; EGF; 1.  
 DR SMART: SM00051; DSL; 1.  
 DR SMART: SM00181; EGF; 2.  
 DR PROSITE: PS00022; EGF\_1; 2.  
 DR PROSITE: PS01186; EGF\_2; 2.  
 DR PROSITE: PS00026; EGF\_3; 2.  
 KM Differentiation; Repeat; Transmembrane; EGF-like domain;  
 KM Glycoprotein; Signal.  
 FT SIGNAL 1 15 POTENTIAL.  
 FT CHAIN 16 402 LAG-2 PROTEIN.

FT DOMAIN 16 279 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 280 306 POTENTIAL.  
 FT DOMAIN 307 402 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 122 166 DSL.  
 FT DOMAIN 171 216 EGF-LIKE 1.  
 FT DOMAIN 229 266 EGF-LIKE 2.  
 FT DISULFID 175 183 BY SIMILARITY.  
 FT DISULFID 177 204 BY SIMILARITY.  
 FT DISULFID 206 215 BY SIMILARITY.  
 FT DISULFID 233 245 BY SIMILARITY.  
 FT DISULFID 239 254 BY SIMILARITY.  
 FT DISULFID 256 265 BY SIMILARITY.  
 FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 105 105 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 194 194 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT MUTAGEN 262 262 G->D: IN SA37; SUPPRESSES LIN-12  
 FUNCTION.

SEQ SEQUENCE 402 AA; 44603 MW; 076A9DA0288A123 CRC64;  
 Query Match 40.4%; Score 44; DB 1; Length 402;  
 Best Local Similarity 41.2%; Pred. No. 13;  
 Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 3 PHPTNKHLYVCESVNG 19  
 Db 193 PNOTSSNEOLICECTNG 209

RESULT 13  
 RS10 TREPA STANDARD; PRT; 102 AA.  
 AC 083218;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DT 30S ribosomal protein S10.  
 GN RPSJ OR TP0188.  
 OS Treponema pallidum.  
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.  
 OX NCBI\_TaxID=160;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=Nichols;  
 RX MEDLINE=9832770; PubMed=9665876;  
 RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,  
 RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,  
 RA Sodegren E., Hardam J.M., McLeod M.P., Salberg S., Peterson J.,  
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Uetshack T.,  
 RA McDonald J., Attiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith R.O.,  
 RA Venter J.C.;  
 RT "Complete genome sequence of Treponema pallidum, the syphilis  
 spirochete";  
 RL Science 281:375-388(1998).  
 CC - FUNCTION: Involved in the binding of tRNA to the ribosomes (By  
 similarity).  
 CC - SIMILARITY: Belongs to the S10P family of ribosomal proteins.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AB001202; AAC65173.1; -.  
 DR PIR; F71354; F71354.  
 DR TIGR; TP0188; -.  
 DR HAMAP; MF00508; -; 1.  
 DR InterPro; IPR001848; Ribosomal\_S10.  
 DR InterPro; IPR005731; Ribosomal\_S10\_B.  
 DR Pfam; PF00338; Ribosomal\_S10; 1.

DR PRINTS; PR00971; RIBOSOMALS10.  
 DR ProDom; PD001272; Ribosomal\_S10; 1.  
 DR TIGRFAMs; TIGR01049; rpsJ\_bact; 1.  
 DR PROSITE; PS00361; RIBOSOMAL\_S10; 1.  
 KW Ribosomal protein; Complete proteome.  
 SQ SEQUENCE 102 AA; 11584 MW; 2C73616A8A01E9E4 CRC64;

Query Match 39.4%; Score 43; DB 1; Length 102;  
 Best Local Similarity 53.3%; Pred. No. 4.1;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 PHPTNKHLYVCES 16  
 Db 40 IPLPTRMKFTVLR 54

RESULT 14  
 VSP3 TRIGA STANDARD; PRT; 258 AA.  
 ID VSP3 TRIGA  
 AC 013063;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DT Venom serine proteinase 3 precursor (EC 3.4.21.-).  
 GN TLG3.  
 OS Trimeresurus gramineus (Indian green tree viper) (Green habu snake).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Lepidosaurs; Squamata; Scleroglossa; Serpentes; Colubroidae;  
 OC Viperidae; Crotalinae; Trimeresurus.  
 OX NCBI\_TaxID=8767;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP TISSUE=Venom gland;  
 RX MEDLINE=97096898; PubMed=8941719;  
 RA Deshmara M., Ogawa T., Nakashima K., Nobuhisa I., Chijiwa T.,  
 RA Shinohegashi Y., Fukumaki Y., Niwa M., Yamashina I., Hattori S.,  
 RA Ohno M.;

RT "Accelerated evolution of crotalinae snake venom gland serine  
 proteases";  
 RL FEBS Lett. 397:83-88(1996).  
 CC - FUNCTION: Thrombin-like snake venom serine protease.  
 CC - SUBCELLULAR LOCATION: Secreted.  
 CC - TISSUE SPECIFICITY: Expressed by the venom gland.  
 CC - SIMILARITY: Belongs to peptidase family S1. Snake venom subfamily.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D67085; BAA19983.1; -.  
 DR HSSP; P00763; IDPO.  
 DR MEROPS; S01.185; -.  
 DR InterPro; IPR009003; Cys Ser trypsin.  
 DR InterPro; IPR001254; Peptidase S1.  
 DR InterPro; IPR001114; Peptidase\_S1A.  
 DR Pfam; PF00089; trypsin.1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR SMART; SM00020; TRYP\_SPC.1.  
 DR PROSITE; PS00240; TRYPsin DOM; 1.  
 DR PROSITE; PS00134; TRYPsin HIS; 1.  
 DR PROSITE; PS00135; TRYPsin SER; 1.  
 KM Hydrolyse; Serine protease; Glycoprotein; Zymogen; Signal.  
 FT SIGNAL 1 18  
 FT PROPEP 19 24  
 FT CHAIN 25 258  
 FT ACT\_SITE 65 65  
 FT ACT\_SITE 110 110  
 FT ACT\_SITE 204 204  
 FT DISULFID 31 163  
 BY SIMILARITY.  
 VENOM SERINE PROTEINASE 3.  
 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 CHARGE RELAY SYSTEM (BY SIMILARITY).



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: March 22, 2004, 06:39:53 ; Search time 3.92324 Seconds  
(without alignments)  
1568.003 Million cell updates/sec

Title: US-09-662-293-11

Perfect score: 109

Sequence: 1 DIPHTNTHKIVCSVNG 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_phage:\*  
9: sp\_plant:\*  
10: sp\_ricent:\*  
11: sp\_virus:\*  
12: sp\_vertebrate:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	103	94.5	555	5 Q9U6R7	Q9U6R7 dermatophag
2	56	51.4	1635	5 Q17412	Q17412 aedes aegypt
3	55	50.5	517	10 Q9S792	Q9S792 nicotiana t
4	55	50.5	517	10 Q9XFT4	Q9XFT4 nicotiana t
5	54	49.5	437	5 Q8T015	Q8T015 drosophila
6	54	49.5	563	5 Q9VST9	Q9VST9 drosophila
7	54	49.5	2382	5 Q9B119	Q9B119 drosophila
8	54	49.5	2786	5 Q9VSU2	Q9VSU2 drosophila
9	53.5	49.1	66	9 Q03967	Q03967 bacteriophag
10	53	48.6	525	5 Q44079	Q44079 anopheles g
11	52	47.7	727	5 Q21139	Q21139 caenorhabdi
12	51	46.8	523	5 Q8MNX4	Q8MNX4 caenorhabdi
13	51	46.8	721	10 Q94DE8	Q94DE8 cyza saati
14	49	45.0	237	5 Q9VR79	Q9VR79 drosophila
15	49	45.0	504	5 Q9NGK8	Q9NGK8 wuchereria
16	49	45.0	647	15 Q9GR89	Q9GR89 human t-lym

17	49	45.0	651	15 Q56227	Q56227 human t-lym
18	49	45.0	796	5 Q9VTR4	Q9VTR4 drosophila
19	49	45.0	796	5 Q8MRG9	Q8MRG9 drosophila
20	49	45.0	1273	15 Q9QRA1	Q9QRA1 human t-lym
21	49	45.0	1462	15 Q56228	Q56228 human t-lym
22	47	43.1	260	13 Q8UXX1	Q8UXX1 agkistrodon
23	47	43.1	290	10 Q9ARD1	Q9ARD1 lycopersico
24	47	43.1	486	5 Q8TSC4	Q8TSC4 aedes aegypt
25	47	43.1	841	12 Q9DMY1	Q9DMY1 human herpe
26	47	43.1	841	12 Q9DX81	Q9DX81 human herpe
27	47	43.1	841	12 Q98VL1	Q98VL1 human herpe
28	47	43.1	841	12 Q8OAC3	Q8OAC3 human herpe
29	47	43.1	841	12 Q8OAC1	Q8OAC1 human herpe
30	47	43.1	852	12 Q66030	Q66030 cercopithe
31	47	43.1	894	5 Q8I511	Q8I511 plasmodium
32	47	43.1	1290	5 Q9VTR8	Q9VTR8 drosophila
33	47	43.1	1324	5 Q8SYK2	Q8SYK2 drosophila
34	46	42.2	183	10 Q7XZC8	Q7XZC8 nicotiana t
35	46	42.2	503	5 Q9NAR8	Q9NAR8 branchiosto
36	46	42.2	520	10 Q43531	Q43531 liliu long
37	46	42.2	1322	5 Q9NTS5	Q9NTS5 anopheles g
38	46	42.2	1322	5 Q9NAT0	Q9NAT0 anopheles g
39	46	42.2	1344	16 Q8NL84	Q8NL84 cornebracte
40	45	41.3	115	12 Q92502	Q92502 bombyx mori
41	45	41.3	118	5 Q27454	Q27454 brugia mala
42	45	41.3	256	5 Q9VTR7	Q9VTR7 drosophila
43	45	41.3	260	13 Q7T229	Q7T229 boehrops ja
44	45	41.3	332	5 Q9VKI8	Q9VKI8 drosophila
45	45	41.3	523	10 Q9ZV69	Q9ZV69 arabidopsis

## ALIGNMENTS

RESULT 1  
ID Q9U6R7 PRELIMINARY; PRT; 555 AA.  
AC Q9U6R7;  
DT 01-MAY-2000 (TREMURel. 13, Created)  
DT 01-MAY-2000 (TREMURel. 13, Last sequence update)  
DT 01-OCT-2003 (TREMURel. 25, Last annotation update)  
DE 98kDa HDM allergen.  
OS Dermatophagoides farinae (House-dust mite).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
OC Acariformes; Sarcoptiformes; Astigmata; Psoroptida; Analgoidea;  
OC Pyroglyphidae; Dermatophagoides.  
OX NCBI\_Taxid=6954;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Weber E.R., Hunter S., Steadman K., McCall C.;  
RT "Cloning and Characterization of a 98 kDa Allergen from  
RT Dermatophagoides farinae."  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF178772; AAD52672.1; -  
DR GO: GO:0005576; C:extracellular; IEA.  
DR GO: GO:0008061; F:chitin binding; IEA.  
DR GO: GO:0016798; F:hydrolase activity; acting on glycosyl bonds; IEA.  
DR GO: GO:0005975; P:carbohydrate metabolism; IEA.  
DR GO: GO:0006030; P:chitin metabolism; IEA.  
DR InterPro: IPR002557; Chitin bind Para.  
DR InterPro: IPR001223; Glyco\_Hydro\_18.  
DR InterPro: IPR001579; Glyco\_Hydro\_18AS.  
DR Pfam: PF00704; Glyco\_Hydro\_18; 1.  
DR ProDom: PD000471; Glyco\_Hydro\_18; 1.  
DR SMART: SM00494; ChEBD2; 1.  
DR SMART: SM00636; Glyco\_18; 1.  
DR PROSITE: PS01095; CHITINASE\_18; 1.  
DR GlycoSite: 555 AA; 63238 MW; OE4564A1A459330B CRC64;  
KW SEQUENCE  
SQ  
Query March 94.5%; Score 103; DB 5; Length 555;  
Best Local Similarity 94.7%; Pred. No. 1e-08;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DHPPTNIHKYLVCESYNG 19  
 DB 512 DHPPTNIHKYLVCESYNG 530

## RESULT 2

017412 PRELIMINARY; PRT; 1635 AA.  
 AC 017412;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Probable chitinase 2 (EC 3.2.1.14).  
 GN CHIT2.  
 OS Aedes aegypti (Yellowfever mosquito).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoides; Aedes.  
 OX NCBI\_TaxId=7159;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98324849; PubMed=9662472;  
 RA de la Vega H., Specht C.A., Liu Y., Robbins P.W.;  
 RT "Chitinases are a multi-gene family in Aedes, Anopheles and  
 Drosophila.".  
 RL Insect Mol. Biol. 7:233-239 (1998).  
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES OF N-  
 ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.  
 CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL  
 CC HYDROLASES).  
 DR EMBL; AF026492; AAB81850.1; -.  
 DR PIR; T14075; T14075.  
 DR GO; GO:0005576; C:extracellular; IEA.  
 DR GO; GO:0008061; F:chitin binding; IEA.  
 DR GO; GO:0008943; F:endochitinase activity; IEA.  
 DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR GO; GO:0006032; P:chitin catabolism; IEA.  
 DR InterPro; IPR002557; Chitin bind. Pera.  
 DR InterPro; IPR001223; Glyco Hydrol. 18.  
 DR InterPro; IPR001579; Glyco\_hydro\_18AS.  
 DR Pfam; PF01607; CBM\_14; 3.  
 DR Pfam; PF00704; Glyco\_hydro\_18; 3.  
 DR ProDom; PD000471; Glyco\_hydro\_18; 3.  
 DR SMART; SMO0494; ChEBD2; 3.  
 DR SMART; SMO0636; Glyco\_18; 3.  
 DR PROSITE; PS01095; CHITINASE\_18; 3.  
 KW Hydrolase; Glycosidase; Chitin degradation; Glycoprotein;  
 KW Multigene family.  
 FT CARBOHYD 192 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 322 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 463 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 749 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 890 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1338 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1479 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1479 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1635 AA; 185993 MW; EA116F83AAC129FA CRC64;

Query Match 51.4%; Score 56; DB 5; Length 1635;  
 Best Local Similarity 50.0%; Pred. No. 2.5;  
 Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 IPHPTNIHKYLVCE 15  
 DB 1134 VPHPTDNCKYITQ 1147

## RESULT 3

09S792 PRELIMINARY; PRT; 517 AA.  
 AC 09S792;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DB Calcium/calmodulin dependent protein kinase.  
 GN CCAMK OR CCAMK-1.  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
 OC Lamiales; Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxId=4097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. XANTHI; TISSUE=Anther;  
 RA Liu Z., Poovath B.W.;  
 RT "Regulated expression of an anther-specific calcium/calmodulin  
 RT dependent protein kinase causes male sterility in plant.";  
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.

GN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. XANTHI; TISSUE=Anther;  
 RA Liu Z.H., Xia M., Poovath B.W.;  
 RT "Chimeric calcium/calmodulin-dependent protein kinase in  
 RT tobacco: differential regulation by calmodulin isoforms.";  
 RL Plant Mol. Biol. 0:0-0 (1998).  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL; U38446; AAF21450.1; -.  
 DR EMBL; AF087813; AAD52092.1; -.  
 DR HSP; G63450; 1A06.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0005509; F:calcium ion binding; IEA.  
 DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.  
 DR InterPro; IPR002048; EF-hand.  
 DR InterPro; IPR000719; Prot\_Kinase.  
 DR InterPro; IPR001125; Recoverin.  
 DR InterPro; IPR002290; Ser\_thr\_Kinase.  
 DR InterPro; IPR008271; Ser\_thr\_Pkin\_AS.  
 DR Pfam; PF00036; ehand; 3.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR00450; RECOVERIN.  
 DR ProDom; PD000012; EF-hand; 1.  
 DR ProDom; PD000001; Prot\_Kinase; 1.  
 DR SMART; SMO0054; Eph; 3.  
 DR SMART; SMO0220; STKc; 1.  
 DR PROSITE; PS00018; EF\_HAND; 3.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.  
 DR PROSITE; PS00108; PROTEIN KINASE ST; 1.  
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
 SQ SEQUENCE 517 AA; 57624 MW; BB670EE29AFB349D CRC64;

Query Match 50.5%; Score 55; DB 10; Length 517;  
 Best Local Similarity 58.8%; Pred. No. 1.1;  
 Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 3 PPHPTNIHKYLVCESYNG 19  
 DB 97 PPHNVHLVDYCEDEPSG 113

## RESULT 4

09XFJ4 PRELIMINARY; PRT; 517 AA.  
 AC 09XFJ4;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Calcium/calmodulin-dependent protein kinase.  
 GN CCAMK OR CCAMK-2.  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
 OC Lamiales; Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxId=4097;



```

RM [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv.XANTH1; TISSUE=anther;
RA Wang W., Liu Z.H., Xia M., Pooviah B.W.;
RT "Chimeric calmodulin-dependent protein kinase in tobacco:
RT differential regulation by calmodulin isoforms.";
RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
RM [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv.XANTH1; TISSUE=anther;
RA Liu Z.H., Xia M., Pooviah B.W.;
RT "Chimeric calmodulin-dependent protein kinase in tobacco:
RT differential regulation by calmodulin isoforms.";
RL Plant Mol. Biol. 0:0-0(1998).
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AF145593; AAD28791.1; -.
DR EMBL; AF145592; AAD28791.1; JOINED.
DR EMBL; U70923; AAD52098.1; -.
DR HSP; Q63450; IAO6.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000719; Recovartin.
DR InterPro; IPR001125; Recovartin.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR008271; Ser_thr_pkin_AS.
DR Pfam; PF00036; ehand; 3.
DR PRINTS; PR00450; RECOVERIN.
DR ProDom; PD000012; EF-hand; 1.
DR ProDom; PD000001; Proc_kinase; 1.
DR SMART; SM00054; Efh; 3.
DR SMART; SM00220; EF_HAND; 1.
DR PROSITE; PS00018; EF_HAND; 3.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR KMP; KMP000001; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 517 AA; 57881 MW; 48F8361E2E80AE61 CRC64;

Query Match 50.5%; Score 55; DB 10; Length 517;
Best Local Similarity 58.8%; Pred. No. 1.1;
Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 3 PHEPTNHYKVCESYNG 19
DB 97 PHEPTNHYKVCESYNG 113

RESULT 5
08T015 PRELIMINARY; PRT; 437 AA.
AC 08T015;
DT 01-UN-2002 (TREMblrel. 21, Created)
DT 01-UN-2002 (TREMblrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE GH28017P.
GN TEQUILA OR CG4821 OR CG4948 OR CG18403.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Gonzalez M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Ceiniker S.;

```

```

RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY069235; AAL9380.1; -.
DR FlyBase; FBgn0023479; Tequila.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0008061; F:chitin binding; IEA.
DR GO; GO:0006030; P:chitin metabolism; IEA.
DR InterPro; IPR002557; Chitin_bind_Pera.
DR Pfam; PF01607; CBM_14; 5.
DR SMART; SM00494; CHeBD2; 5.
SQ SEQUENCE 437 AA; 49481 MW; 1D2D73B1E7E5CF20 CRC64;

Query Match 49.5%; Score 54; DB 5; Length 437;
Best Local Similarity 66.7%; Pred. No. 1.4;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 PHEPTNHYKVC 14
DB 98 PHEPTNHYKVC 109

RESULT 6
08VST9 PRELIMINARY; PRT; 563 AA.
AC 08VST9;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE CG4821 protein.
GN TEQUILA OR CG4821 OR CG4948 OR CG18403.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Avril J.F., Agbayani A., An H.-U., Andrews-Piankovich C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo K., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunker B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Huck C.,
RA Hestlin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagham C.,
RA Jatali M., Kalush F., Karen G.H., Ke Z., Kienison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Meklov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasarman D.A., Weinstein G.M., Weisenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

```

RA Gibbs R.A., Myers B.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*."  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Celinker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Goez J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banton J., An H., Baldwin D., Banton J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Cener A., Champagne M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorett V., Doup L.E., Doyle C., Drenek D., Farfan D.,  
 RA Ferreira S., Frise E., Galle R.F., Gary N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegun C., Jaitai M., Kruse D., Li P., Matti B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunco J.,  
 RA Pacled V., Paragas V., Park S., Patel S., Pfeiffer B., Scheeler F.,  
 RA Phouenavong S., Pittman G.S., Puri V., Richards S., Scheller F.,  
 RA Stapleton M., Strong R., Svitskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of *Drosophila melanogaster* genome."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Miera S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celinker S.E.,  
 RA Clamp W., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Kronmiller B., Marshall B., Milburn G., Richter J., Russo S.,  
 RA Searle S.M., Smith E., Shu S., Smutnak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.D., Lewis S.E.;  
 RT "Annotation of *Drosophila melanogaster* genome."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL EMBL: AEO03553; AAF0322.2; -  
 DR FlyBase: FBgn0023479; Tegula.  
 DR GO: GO:0005576; C:extracellular; IEA.  
 DR GO: GO:0008061; F:chitin binding; IEA.  
 DR GO: GO:0006030; P:chitin metabolism; IEA.  
 DR InterPro: IPR002557; Chitin\_bind\_Pera.  
 DR Pfam: PF01607; CBM\_14; 7.  
 DR SMART: SM00494; CHTBD2; 7.  
 DR SIGNAL  
 SQ SEQUENCE 563 AA; 61963 MW; 9AC106B0F4913F93 CRC64;  
 QY 3 PHPTNKHLYVC 14  
 Db 224 PHPDVHKYLRG 235  
 Query Match 49.5%; Score 54; DB 5; Length 563;  
 Best Local Similarity 66.7%; Pred. No. 1.8;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

RP SEQUENCE FROM N.A.  
 RA Munier A.I., Medzhitov R., Janeway C.A., Lanot R., Zachary D.,  
 RA Capovilla M., Lagaux M.;  
 RT "Gral a *Drosophila* gene coding for several mosaic serine proteases."  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 DR EMBL: AJ309005; CAC35209.1; -  
 DR HSP: P00750; IRTF.  
 DR FlyBase: FBgn0023479; Tegula.  
 DR GO: GO:0005576; C:extracellular; IEA.  
 DR GO: GO:0016020; C:membrane; IEA.  
 DR GO: GO:0008061; F:chitin binding; IEA.  
 DR GO: GO:0004263; F:chymotrypsin activity; IEA.  
 DR GO: GO:0008233; F:peptidase activity; IEA.  
 DR GO: GO:0005044; F:scavenger receptor activity; IEA.  
 DR GO: GO:0004295; F:trypsin activity; IEA.  
 DR GO: GO:0006030; F:chitin metabolism; IEA.  
 DR GO: GO:0006508; F:proteolysis and peptidolysis; IEA.  
 DR InterPro: IPR002557; Chitin\_bind\_Pera.  
 DR InterPro: IPR009003; Cys Ser trypsin.  
 DR InterPro: IPR002172; IDL\_receptor\_A.  
 DR InterPro: IPR001254; Peptidase\_S1.  
 DR InterPro: IPR001314; Peptidase\_S1A.  
 DR Pfam: PF01607; CBM\_14; 15.  
 DR Pfam: PF00057; Idl\_recept\_a; 2.  
 DR Pfam: PF00530; SRCR\_2.  
 DR Pfam: PF00089; trypsin\_1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR PRINTS: PR00261; LDLRECEPTOR.  
 DR PRINTS: PR00258; SPERACTRCPTR.  
 DR SMART: SM00494; CHTBD2; 15.  
 DR SMART: SM00182; LDLa; 2.  
 DR SMART: SM00202; SR; 2.  
 DR PROSITE: PS01209; IDLRA\_1; 1.  
 DR PROSITE: PS0068; IDLRA\_2; 2.  
 DR PROSITE: PS00420; SRCR\_1; 2.  
 DR PROSITE: PS00287; SRCR\_2; 2.  
 DR PROSITE: PS0240; TRYPsin DOM; 1.  
 DR PROSITE: PS00134; TRYPsin HIS; 1.  
 DR PROSITE: PS00135; TRYPsin SER; 1.  
 KM Hydrolase; Protease; Serine protease; Signal.  
 FT SIGNAL  
 SQ SEQUENCE 2382 AA; 264348 MW; 51C85282B06833D4 CRC64;  
 QY 3 PHPTNKHLYVC 14  
 Db 224 PHPDVHKYLRG 235  
 Query Match 49.5%; Score 54; DB 5; Length 2382;  
 Best Local Similarity 66.7%; Pred. No. 8.2;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

RESULT 7  
 Q9B119 PRELIMINARY; PRT; 2382 AA.  
 ID Q9B119;  
 AC Q9B119;  
 DT 01-JUN-2001 (TREMELrel. 17, Created)  
 DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)  
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
 DE TEQUILA OR GRAAL OR CG4821 OR CG4948 OR CG18403.  
 GN TEQUILA OR GRAAL OR CG4821 OR CG4948 OR CG18403.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; *Drosophila*.  
 OC NCBI\_TaxID=7227;  
 RN [1]

RESULT 8  
 Q9VSU2 PRELIMINARY; PRT; 2786 AA.  
 ID Q9VSU2;  
 AC Q9VSU2;  
 DT 01-MAY-2000 (TREMELrel. 13, Created)  
 DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
 DE TEQUILA OR CG4821 OR CG4948 OR CG18403.  
 GN TEQUILA OR CG4821 OR CG4948 OR CG18403.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; *Drosophila*.  
 OC NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amaratunga P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers J.-H.C., Blaise R.G., Chame M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baker E.G., Helt G., Nelson C.R., McKillop G.L.G.,  
 RA Abril J.F., Abayaratne A., An H.-U., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintsov S.,  
 RA Borokova D., Botchan P.R., Bouck P., Brockstein P., Brotler P.,  
 RA Burris K.C., Bussey D.A., Butler H., Cadieu C., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durkin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasner K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,  
 RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laspo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Palzer K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Sidenkiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spirdling A.C., Stepieton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Weissman D.A., Weissbach G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye T., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao X., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195 (2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amaratunga P.G., Brandon R.C., Rogers Y.,  
 RA Barton J., An H., Baldwin D., Barton J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Chame M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Fartan D.,  
 RA Ferriera S., Frise E., Galie R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hoslin D., Howland T.J.,  
 RA Ibegam C., Jatali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuno J.,  
 RA Paclab J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of *Drosophila melanogaster* genome.";  
 RL Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Mistra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
 RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,  
 RA Seale S.W.J., Smith E., Shu S., Smutnick F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of *Drosophila melanogaster* genome.";  
 RL Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;  
 RL Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Flybase;

RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.  
 DR EMBL: AE003553; AF50319.3; -.  
 DR HSPB; P00750; IRTF.  
 DR FLYBASE: FBgn0023479; Tegula.  
 DR GO: GO:0005576; C:extracellular; IEA.  
 DR GO: GO:0016020; C:membrane; IEA.  
 DR GO: GO:0005524; F:ATP binding; IEA.  
 DR GO: GO:0008061; F:chitin binding; IEA.  
 DR GO: GO:0004263; F:chymotrypsin activity; IEA.  
 DR GO: GO:0008233; F:peptidase activity; IEA.  
 DR GO: GO:0005044; F:scavenger receptor activity; IEA.  
 DR GO: GO:0004295; F:trypsin activity; IEA.  
 DR GO: GO:0006030; F:chitin metabolism; IEA.  
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO: GO:0015992; P:proton transport; IEA.  
 DR InterPro: IPR00194; ATPase a/bcentre.  
 DR InterPro: IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro: IPR02172; LDL\_receptor\_A.  
 DR InterPro: IPR001254; peptidase\_S1.  
 DR InterPro: IPR001314; peptidase\_S1A.  
 DR InterPro: IPR001190; Srcr\_receptor.  
 DR Pfam: PF01607; CBM\_14; 15\_receptor.  
 DR Pfam: PF00057; 1d1\_recept\_a; 2.  
 DR Pfam: PF00530; SRCR; 2.  
 DR Pfam: PF00089; trypsin; 1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR PRINTS: PR00261; LDLRECEPTOR.  
 DR PRINTS: PR00258; SPERACTRPT.  
 DR SMART: SM00494; CUBD2; 15.  
 DR SMART: SM00192; LDLB; 2.  
 DR SMART: SM00202; SR; 2.  
 DR SMART: SM00020; TRYPSIN\_SPC; 1.  
 DR PROSITE: PS00152; ATPASE\_ALPHA\_BETA; 1.  
 DR PROSITE: PS01209; LDLRA\_1; 1.  
 DR PROSITE: PS00068; LDLRA\_2; 2.  
 DR PROSITE: PS00420; SRCR\_1; 2.  
 DR PROSITE: PS00287; SRCR\_2; 2.  
 DR PROSITE: PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
 DR Hydrobase: Protease; Serine protease.  
 SQ SEQUENCE 2786 AA; 308362 MW; 2BED7A7DA50020C76 CRC64;  
 Query Match 49.5%; Score 54; DB 5; Length 2786;  
 Best Local Similarity 66.7%; Pred. No. 9.7;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 3 PHPTNHYKYLVC 14  
 DB 224 PPHVDVHKYDRC 235  
 ID 003967 PRELIMINARY; PRT; 66 AA.  
 AC 003967;  
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE EXCISIONASE.  
 GN XIS OR GP66.  
 OS Bacteriophage phi16.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.  
 OX NCBI\_TaxID=52979;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97225795; PubMed=9073065;  
 RA Kodaira K.I., Oki M., Kakikawa M., Watanabe N., Hirakawa M.,  
 RA Yamada K., Takeo A.;  
 RT "Genome structure of the *Lactococcus lactis* temperate phage phi 16: the whole genome sequence and the putative promoter/repressor system.";

```

RL Gene 187:45-53(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Kodelira K.;
RL Submitted (JUG-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X98106; CAA6757.1; -.
DR EMBL; X90510; CAA62091.1; -.
SQ SEQUENCE 66 AA; 7557 MW; 4FDA25D9440CD1E4 CRC64;

Query Match 49.1%; Score 53.5; DB 9; Length 66;
Best Local Similarity 50.0%; Pred. No. 0.22;
Matches 9; Conservative 4; Mismatches 2; Indels 3; Gaps 1;

Qy 1 DIPPTNI--HKYLVC 15
Db 37 DVPPPTNIASHYHFLACD 54

RESULT 10
ID 044079 PRELIMINARY; PRT; 525 AA.
AC 044079;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Chitinase.
GN AGCHT-1.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxId=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Gut;
RX MEDLINE=98030563; PubMed=9360958;
RA Shen Z., Jacobs-Lorena M.;
RT "Characterization of a novel gut-specific chitinase gene from the
RT human malaria vector Anopheles gambiae.";
RL J. Biol. Chem. 272:28895-28900 (1997).
DR EMBL; AF008575; AAB8764.1; -.
DR PIR; T44445; T44445.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0008061; F:chitin binding; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0006030; P:chitin metabolism; IEA.
DR InterPro; IPR002557; Chitin bind. Perz.
DR InterPro; IPR001223; Glyco_hydro_18.
DR InterPro; IPR001579; Glyco_hydro_18AS.
DR Pfam; PF01607; CBM_14; 1.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR ProDom; PD000471; Glyco_hydro_18; 1.
DR SMART; SM00494; ChcBD2; 1.
DR SMART; SM00636; Glyco_18; 1.
DR PROSITE; PS01095; CHITINASE_18; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 525 AA; 57211 MW; 3234360EEFF36165 CRC64;

Query Match 48.6%; Score 53; DB 5; Length 525;
Best Local Similarity 53.8%; Pred. No. 2.4;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 2 IPHPTNIHKYLVC 14
Db 478 VPHPTNHCARYIC 490

RESULT 11
ID 021139 PRELIMINARY; PRT; 727 AA.
AC 021139;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)

```

```

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Synaplogenesi abnormal protein 1.
GN K02E10.8 OR SYG-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Wilson R.;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wu X., Leimbach D.;
RT "The sequence of C. elegans cosmid K02E10.";
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (May-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; U40942; AAC47074.5; -.
DR PIR; T16525; T16525.
DR WormRep; K02E10.8; CE33930.
DR InterPro; IPR003599; IG_1.
DR InterPro; IPR007110; IG_1-like.
DR InterPro; IPR003598; IG_c2.
DR Pfam; PF00047; IG; 4.
DR SMART; SM00409; IG; 4.
DR SMART; SM00408; IGc2; 4.
DR PROSITE; PS00835; IG_LIKE; 5.
SQ SEQUENCE 727 AA; 80863 MW; C623B14C92205E2A CRC64;

Query Match 47.7%; Score 52; DB 5; Length 727;
Best Local Similarity 60.0%; Pred. No. 5.1;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 IPHPTNIHKYLVCES 16
Db 234 IPRPDHKLICIS 248

RESULT 12
ID 08NMK4 PRELIMINARY; PRT; 523 AA.
AC 08NMK4;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN T11F1.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Waterston R.;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R., Mameley P.;
RT "The sequence of C. elegans cosmid T11F1.";

```



```

RA Ciamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak P., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,
RT "Annotation of Drosophila melanogaster genome."
RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.,
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunco J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouphenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RA EMBL; AEO03572; AAF50927.2; -.
DR EMBL; BT001512; AAT71267.1; -.
DR FlyBase; FBgn0031097; CG17052.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0008061; F:chitin binding; IEA.
DR GO; GO:0006030; P:chitin metabolism; IEA.
DR InterPro; IPR002557; Chitin_bind_Pera.
DR Pfam; PF01607; CBM_14; 3.
DR SMART; SM00494; ChtBD2; 3.
SQ SEQUENCE 237 AA; 26564 MW; D477AE83FD3B91A7 CRC64;

Query Match 45.0%; Score 49; DB 5; Length 237;
Best Local Similarity 52.9%; Pred. No. 4.9;
Matches 9; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 3 PHPTNKHVLCESVNG 19
DB 181 PHPTDCKFYVC--LNG 195

RESULT 15
Q9NGK8 PRELIMINARY; PRT; 504 AA.
ID Q9NGK8;
AC Q9NGK8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cuticular endochitinase.
OS Wuchereria bancrofti.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Wuchereria.
OX NCBI_TaxID=6293;
RN [1]
RP SEQUENCE FROM N.A.
RA Varatharajulu V., Kubofcik J., Williams S.A., Kallraj P.,
RA Jayaraman K., Nutman T.;
RT "Molecular characterization of endochitinase from Wuchereria bancrofti
RT Microfilial cDNA library."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF250997; AAF66988.1; -.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0008061; F:chitin binding; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0006030; P:chitin metabolism; IEA.
DR InterPro; IPR002557; Chitin_bind_Pera.
DR InterPro; IPR001223; Glyco_hydro_18.
DR InterPro; IPR001579; Glyco_hydro_18A5.
DR Pfam; PF01607; CBM_14; 1.

```

```

DR Pfam; PF00704; Glyco_hydro_18; 1.
DR ProDom; PD000471; Glyco_hydro_18; 1.
DR SMART; SM00494; ChtBD2; 1.
DR SMART; SM00636; Glyco_18; 1.
DR PROSITE; PS01095; CHITINASE_18; 1.
SQ SEQUENCE 504 AA; 56480 MW; D65C888300DE2C27 CRC64;

Query Match 45.0%; Score 49; DB 5; Length 504;
Best Local Similarity 43.8%; Pred. No. 11;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 3 PHPTNKHVLCESVN 18
DB 459 PHPTDCHLFIICGNSN 474

Search completed: March 22, 2004, 06:59:22
Job time : 4.82324 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OK protein - protein search, using sw model

Run on: March 22, 2004, 06:30:23 ; Search time 5.7771 Seconds

(without alignments)  
929.256 Million cell updates/sec

Title: US-09-662-293-11

Perfect score: 109  
Sequence: 1 DIPHPNTNHYKLVCSVNG 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

1: geneseqp1990s:\*\n2: geneseqp1990s:\*\n3: geneseqp2000s:\*\n4: geneseqp2001s:\*\n5: geneseqp2002s:\*\n6: geneseqp2003as:\*\n7: geneseqp2003bs:\*\n8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	109	100.0	19	AAV52520	AAV52520 House dus
2	109	100.0	19	AAU96324	AAU96324 Der HMW-m
3	109	100.0	20	AAV52512	AAV52512 House dus
4	109	100.0	20	AAU96316	AAU96316 Der HMW-m
5	103	94.5	536	AAV52525	AAV52525 House dus
6	103	94.5	536	AAU96329	AAU96329 Der HMW-m
7	103	94.5	555	AAV52523	AAV52523 House dus
8	103	94.5	555	AAU96327	AAU96327 Der HMW-m
9	103	94.5	555	AAU96328	AAU96328 Der HMW-m
10	76.5	70.2	490	AAV52535	AAV52535 D. pteron
11	76.5	70.2	490	AAU96339	AAU96339 Der HMW-m
12	76.5	70.2	509	AAV52533	AAV52533 D. pteron
13	76.5	70.2	509	AAU96337	AAU96337 Der HMW-m
14	76.5	70.2	509	AAU96338	AAU96338 Der HMW-m
15	67	61.5	143	AAO17674	AAO17674 B tropica
16	67	61.5	143	AAO17673	AAO17673 B tropica
17	59	54.1	67	ABP33819	ABP33819 Human ORF
18	55	50.5	517	AAW30919	AAW30919 Tobacco c
19	54	49.5	516	ABB61121	ABB61121 Drosophi
20	53	48.6	525	ABP72636	ABP72636 Anopheles
21	49	45.0	248	ABB66151	ABB66151 Drosophi
22	49	45.0	796	ABB63128	ABB63128 Drosophi
23	47	43.1	291	ABBA7881	ABBA7881 Listeria
24	47	43.1	291	ABP97893	ABP97893 Sycopan 3
25	47	43.1	362	ABG13234	ABG13234 Novel hum

26	47	43.1	841	1	AAV70045	AAV70045 Varicella
27	47	43.1	1290	4	ABB62818	ABB62818 Drosophi
28	46	42.2	520	2	AAW30918	AAW30918 Lilly calc
29	46	42.2	874	4	ABG324600	ABG324600 Novel hum
30	46	42.2	1344	4	AAG33134	AAG33134 C glutam
31	45	41.3	57	1	AAV90413	AAV90413 Plasmodiu
32	45	41.3	80	5	AAU79410	AAU79410 Human 8.8
33	45	41.3	110	2	AAW15402	AAW15402 BmPV TYP
34	45	41.3	268	4	ABB69017	ABB69017 Drosophi
35	45	41.3	332	4	ABB62228	ABB62228 Drosophi
36	45	41.3	393	3	AAG31918	AAG31918 Arabidops
37	45	41.3	454	3	AAG31917	AAG31917 Arabidops
38	45	41.3	458	4	ABB66690	ABB66690 Drosophi
39	45	41.3	504	6	ABP72620	ABP72620 Brugia ma
40	45	41.3	523	3	AAG31916	AAG31916 Arabidops
41	45	41.3	592	7	ADB86252	ADB86252 Hedgehog
42	45	41.3	1077	4	ABB61013	ABB61013 Drosophi
43	45	41.3	2016	4	ABB66424	ABB66424 Drosophi
44	45	41.3	2016	6	ABG73274	ABG73274 D. melano
45	44.5	40.8	835	4	ABB58693	ABB58693 Drosophi

#### ALIGNMENTS

##### RESULT 1

AAV52520 standard; peptide; 19 AA.

AAV52520;

22-FEB-2000 (first entry)

House dust mite allergen protein (map) A/B fragment map(10).

Mite allergen protein; map; high molecular weight; HMW-map; allergy;

house dust mite; IGE; immunoglobulin E; allergen; mapA; mapB;

hypersensitivity reaction; therapy; diagnosis; human; feline;

canine; veterinary; antibody; vaccine; immunisation.

OS Dermatophagoides farinae.

PN WO9954349-A2.

PD 28-OCT-1999.

PF 16-APR-1999; 99WO-US006524.

PR 17-APR-1998; 98US-00062013.

PI 13-MAY-1998; 98US-0085295P.

PS 02-SEP-1998; 98US-0098909P.

PA (HRSK-) HESKA CORP.

PI McCall CA, Hunter SW, Weber ER;

WP; 2000-052700/04.

Novel high molecular weight Dermatophagoides nucleic acid polypeptides

used to modify an animal's hypersensitivity to mite allergens.

Claim 3; Page 70; 154pp; English.

Sequences AAV52510-Y52522 represent proteolytic fragments of

Dermatophagoides farinae high molecular weight mite allergen protein (HMW

map) composition. The HMW-map composition was isolated from a D. farinae

homogenate by gel filtration, with each fraction being analysed for the

presence of proteins that bound to IGE present in mite-allergic dog

antisera. The HMW-map composition comprises mapA (a 109 kD protein) and

mapB (98 kD). Mite allergenic proteins and peptides, and nucleic acids

encoding them, may be used in therapeutic compositions to modify an

animal's hypersensitivity reaction to mite allergens. Animals that may be

treated include mammals and birds, especially felines, canines, equines,

humans, other pets, and work or domestic animals. The proteins or fragments may also be used to diagnose allergies via a skin test. The proteins and peptides can also be used to raise antibodies, which have a variety of potential uses. For example, they can be used as vaccines to passively immunise animals against dust mite hypersensitivity, as positive controls in test kits and as tools to recover desired dust mite allergens from a mixture of proteins

XX Sequence 19 AA;

Query Match 100.0%; Score 109; DB 3; Length 19;

Best Local Similarity 100.0%; Pred. No. 2.6e-11; Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DIPHPTNHHKYLVCESVNG 19  
DB 1 DIPHPTNHHKYLVCESVNG 19

RESULT 2  
AAU96324  
ID AAU96324 standard; peptide; 19 AA.

XX AAU96324;

DT 15-JUL-2002 (first entry)

DE Der HMW-map polypeptide #11.

XX Der HMW-map: American house dust mite; antiallergic; mite; IGB;  
KM mite allergenic protein; immunoglobulin E; hypersensitivity;  
XX immunocomplex formation.

XX Dermatophagoides farinae.

XX WC00222807-A2.

XX 21-MAR-2002.

XX 14-SEP-2001; 2001MO-US028730.

XX 14-SEP-2000; 2000US-00662293.

XX (HESK-) HESKA CORP.

PI McCall CA, Hunter SW, Weber ER;

XX WPI; 2002-351888/38.

PT New mite allergenic protein isolated from Dermatophagoides, designated  
PT Der HMW-map protein, useful as a vaccine for treating mite allergy.

XX Claim 12; Page 71; 161pp; English.

XX The invention relates to an isolated mite allergenic protein of  
CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic  
CC acid. The Der HMW-map protein is useful for eliciting an immune response  
CC against Der HMW-map protein. The protein or a reagent comprising a non-  
CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
CC cat) susceptible to or having an allergic response to a mite. A  
CC therapeutic composition is useful for desensitising a host animal to an  
CC allergic response to a mite. The DNA and protein can be used in the  
CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
CC of immunoglobulin (Ig)E or Der HMW-map protein activity associated with a  
CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
CC reducing hypersensitivity responses to mite allergens, and as vaccines  
CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
CC represent Der HMW-map polypeptides of the invention

XX Sequence 19 AA;

Query Match 100.0%; Score 109; DB 5; Length 19;

Best Local Similarity 100.0%; Pred. No. 2.6e-11; Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DIPHPTNHHKYLVCESVNG 19  
DB 1 DIPHPTNHHKYLVCESVNG 19

RESULT 3

AAU96312  
ID AAU96312 standard; peptide; 20 AA.

XX AAU96312;

DT 22-FEB-2000 (first entry)

DE House dust mite allergen protein (map) A/B fragment map(2).

XX Mite allergen protein; map; high molecular weight; HMW-map; allergy;  
KM house dust mite; IGB; immunoglobulin E; allergen; mapA; mapB;  
KM hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
KM canine; veterinary; antibody; vaccine; immunisation.

XX Dermatophagoides farinae.

XX WO954349-A2.

PD 28-OCT-1999.

XX 16-APR-1999; 99WO-US008524.

XX 17-APR-1998; 98US-00062013.

XX 13-MAY-1998; 98US-0085235P.

XX 02-SEP-1998; 98US-0098909P.

XX (HESK-) HESKA CORP.

PI McCall CA, Hunter SW, Weber ER;

XX WPI; 2000-052700/04.

PT Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
PT used to modify an animals' hypersensitivity to mite allergens.

XX Claim 3; Page 69; 154pp; English.

XX Sequences AAU96310-AAU96322 represent proteolytic fragments of  
CC Dermatophagoides farinae high molecular weight mite allergen protein (HMW  
CC -map) composition. The HMW-map composition was isolated from a D. farinae  
CC homogenate by gel filtration, with each fraction being analysed for the  
CC presence of proteins that bound to IgE present in mite-allergic dog  
CC antisera. The HMW-map composition comprises mapA (a 109 kD protein) and  
CC mapB (98 kD). Mite allergenic proteins and peptides, and nucleic acids  
CC encoding them, may be used in therapeutic compositions to modify an  
CC animal's hypersensitivity reaction to mite allergens. Animals that may be  
CC treated include mammals and birds, especially felines, canines, equines,  
CC humans, other pets, and work or domestic animals. The proteins or  
CC fragments may also be used to diagnose allergies via a skin test. The  
CC proteins and peptides can also be used to raise antibodies, which have a  
CC variety of potential uses. For example, they can be used as vaccines to  
CC passively immunise animals against dust mite hypersensitivity, as  
CC positive controls in test kits and as tools to recover desired dust mite  
CC allergens from a mixture of proteins

XX Sequence 20 AA;

Query Match 100.0%; Score 109; DB 3; Length 20;

Best Local Similarity 100.0%; Pred. No. 2.8e-11; Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DIPHPTNHHKYLVCESVNG 19  
DB 1 DIPHPTNHHKYLVCESVNG 19



ID	AAU96316 standard; peptide; 20 AA.
AC	AAU96316;
DT	15-JUL-2002 (first entry)
DE	Der HMW-map polypeptide #3.
KM	Der HMW-map; American house dust mite; anti-allergic; mite; IgE, mite allergenic protein; immunoglobulin E; hypersensitivity; immunocomplex formation.
OS	Dermatophagoides farinae.
PN	WO200222807-A2.
PD	21-MAR-2002.
PF	14-SEP-2001; 2001WO-US028730.
PR	14-SEP-2000; 2000US-00662293.
PA	(HESK-) HESKA CORP.
PI	Mccall CA, Hunter SW, Weber ER,
DR	WPI; 2002-351888/38.
PT	New mite allergenic protein isolated from Dermatophagoides, designated Der HMW-map protein, useful as a vaccine for treating mite allergy.  Claim 12; Page 70; 16pp; English.
PS	The invention relates to an isolated mite allergenic protein of Dermatophagoides, designated Der HMW-map protein, and its related nucleic acid. The Der HMW-map protein is useful for eliciting an immune response against Der HMW-map protein. The protein or a reagent comprising a non- proteinaceous epitope is useful for identifying an animal (e.g., dog, cat) susceptible to or having an allergic response to a mite. A therapeutic composition is useful for desensitising a host animal to an allergic response to a mite. The DNA and protein can be used in the detection of anti-Der HMW-map antibodies in animal fluids, and inhibition of immunoglobulin (Ig)E or Der HMW-map protein activity associated with a disease. Antibodies that bind to Der HMW-map are useful for inhibiting binding of proteins to IgE, to prevent immunocomplex formation, thus reducing hypersensitivity responses to mite allergens, and as vaccines against mite allergen hypersensitivity. Sequences AAU96314-AAU96342 represent Der HMW-map polypeptides of the invention
CC	Sequence 20 AA:
SQ	
OY	Query Match            100.0%; Score 109; DB 5; Length 20; Best Local Similarity   100.0%; Fred.No. 2.8e-11; Matches     19; Conservative      0; Mismatches       0; Indels          0; Gaps            0;
DB	1 DIPHNTIHKYLVCSEVSG 19   1 DIPHNTIHKYLVCSEVSG 19
RESULT 5	
ID	AAU96316 standard; protein; 536 AA.
XX	AAU96316;
DT	22-FEB-2000 (first entry)
DE	House dust mite (D. farinae) mite allergen protein (map) FDefF98-536.

XX Mite allergen protein; map; high molecular weight; HMM-map; allergy;  
 KW house dust mite; IGE; immunoglobulin E; allergen; MAPA; mAPB;  
 KW hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
 KW canine; veterinary; antibody; vaccine; immunisation.  
 XX  
 XX  
 OS Dermatophagoides farinae.  
 XX  
 XX  
 FN W09954349-A2.  
 PD 28-OCT-1999.  
 XX  
 PE 16-APR-1999; 99WO-US008524.  
 XX  
 PR 17-APR-1998; 98US-00062013.  
 PR 13-MAY-1998; 98US-0085295P.  
 PR 02-SEP-1998; 98US-0098909P.  
 XX  
 PA (HESK-) HESKA CORP.  
 PI McCall CA, Hunter SM, Weber ER;  
 DR N-PEDB; AAZ38579, AAZ38580.  
 XX  
 PT Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
 XX used to modify an animals' hypersensitivity to mite allergens.  
 XX  
 PS Claim 3, Page 125-127, 154pp; English.  
 XX  
 XX This sequence represents Dermatophagoides farinae mite allergen protein  
 CC (map) Pderf98-536, the mature form of pderf98-555 (AAV52523). Pderf98-536  
 CC has a molecular weight of 98 kD, comprising 536 amino acids, and is a  
 CC component of the Dermatophagoides farinae high molecular weight mite  
 CC allergen protein (HMM-map) composition. The HMM-map composition was  
 CC isolated from a D. farinae homogenate by gel filtration, with each  
 CC fraction being analysed for the presence of proteins that bound to IGE  
 CC present in mite-allergic dog antisera. Mite allergenic proteins and  
 CC peptides, and nucleic acids encoding them, may be used in therapeutic  
 CC compositions to modify an animal's hypersensitivity reaction to mite  
 CC allergens. Animals that may be treated include mammals and birds,  
 CC especially felines, canines, equines, humans, other pets, and dogs or  
 CC domestic animals. The proteins or fragments may also be used to diagnose  
 CC allergies via a skin test. The proteins and peptides can also be used to  
 CC raise antibodies, which have a variety of potential uses. For example,  
 CC they can be used as vaccines to passively immunise animals against dust  
 CC mite hypersensitivity, as positive controls in test kits and as tools to  
 CC recover desired dust mite allergens from a mixture of proteins  
 XX  
 SQ Sequence 536 AA;  
 QY  
 Query Match 94.5%; Score 103; DB 3; Length 536;  
 Best Local Similarity 94.7%; Pred. No. 1.3e-08;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0  
 1 DIPHPTNIHKYLVCSYVNG 19  
 |||||  
 493 DIPHPTNIHKYLVCSYVNG 511  
 RESULT 6  
 AAU96329  
 ID AAU96329 standard; protein; 536 AA.  
 XX  
 AC AAU96329;  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX  
 DE Der HMM-map polypeptide #16.  
 XX  
 KW Der HMM-map; American house dust mite; anti-allergic; mite; IGE;  
 KW mite allergenic protein; immunoglobulin E; hypersensitivity;  
 KW immunocomplex formation.

XX OS Dermatophagoides farinae.  
 XX PN WO200222807-A2.  
 XX PD 21-MAR-2002.  
 XX PF 14-SEP-2001; 2001WO-US028730.  
 XX PR 14-SEP-2000; 2000US-00662293.  
 XX PA (HESK-) HESKA CORP.  
 XX PI Mccall CA, Hunter SW, Weber ER;  
 XX DR WPI; 2002-351888/38.  
 XX DR N-PSDB; ABK69575.  
 XX PT New mite allergenic protein isolated from Dermatophagoides, designated  
 XX PT Der HMW-map protein, useful as a vaccine for treating mite allergy.  
 XX PS Claim 12; Page 125-127; 161pp; English.  
 CC The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic  
 CC acid. The Der HMW-map protein is useful for eliciting an immune response  
 CC against Der HMW-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitizing a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (IgE) or Der HMW-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HMW-map polypeptides of the invention  
 XX SQ Sequence 536 AA;  
 SQ Query Match 94.5%; Score 103; DB 5; Length 536;  
 Best Local Similarity 94.7%; Pred. No. 1.3e-08;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 DIPHPNTHKYLVCESVNG 19  
 Db 493 DIPHPNTHKYLVCESVNG 511  
 RESULT 7  
 AAU95253  
 ID AAU95253 standard; protein; 555 AA.  
 XX AC AAU95253;  
 XX DT 22-FEB-2000 (first entry)  
 XX DE House dust mite (D. farinae) mite allergen protein (map) PDerf98-555.  
 XX DE Mite allergen protein; map; high molecular weight; HMW-map; allergy;  
 KM house dust mite; IgE; immunoglobulin E; allergen; mapA; mapB;  
 KM hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
 KM canine; veterinary; antibody; vaccine; immunisation.  
 XX OS Dermatophagoides farinae.  
 XX PH Key Location/Qualifiers  
 FT Peptide 1..19  
 FT /note= "Signal peptide"  
 FT Protein 20..555  
 FT /note= "Mature PDerf98-555"

PN WO9954349-A2.  
 XX PD 28-OCT-1999.  
 XX PF 16-APR-1999; 99WO-US008524.  
 XX PR 17-APR-1998; 98US-00062013.  
 XX PR 13-MAY-1998; 98US-0085225P.  
 XX PR 02-SEP-1998; 98US-0098909P.  
 XX PA (HESK-) HESKA CORP.  
 XX PI Mccall CA, Hunter SW, Weber ER;  
 XX DR WPI; 2000-052700/04.  
 XX DR N-PSDB; AA238575, AA238576, AA238577, AA238578.  
 XX PT Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
 XX PT used to modify an animals' hypersensitivity to mite allergens.  
 XX PS Claim 3; Page 111-113; 154pp; English.  
 CC This sequence represents Dermatophagoides farinae mite allergen protein  
 CC (map) PDerf98-555. PDerf98-555 has a molecular weight of 98 kD,  
 CC comprising 555 amino acids, and is a component of the Dermatophagoides  
 CC farinae high molecular weight mite allergen protein (HMW-map)  
 CC composition. The HMW-map composition was isolated from a D. farinae  
 CC homogenate by gel filtration, with each fraction being analysed for the  
 CC presence of proteins that bound to IgE present in mite-allergic dog  
 CC encubers. Mite allergenic proteins and peptides, and nucleic acids  
 CC encoding them, may be used in therapeutic compositions to modify an  
 CC animal's hypersensitivity reaction to mite allergens. Animals that may be  
 CC treated include mammals and birds, especially felines, canines, equines,  
 CC humans, other pets, and work or domestic animals. The proteins or  
 CC fragments may also be used to diagnose allergies via a skin test. The  
 CC proteins and peptides can also be used to raise antibodies, which have a  
 CC variety of potential uses. For example, they can be used as vaccines to  
 CC passively immunise animals against dust mite hypersensitivity, as  
 CC positive controls in test kits and as tools to recover desired dust mite  
 CC allergens from a mixture of proteins  
 XX SQ Sequence 555 AA;  
 SQ Query Match 94.5%; Score 103; DB 3; Length 555;  
 Best Local Similarity 94.7%; Pred. No. 1.3e-08;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 DIPHPNTHKYLVCESVNG 19  
 Db 512 DIPHPNTHKYLVCESVNG 530  
 RESULT 8  
 AAU96327  
 ID AAU96327 standard; protein; 555 AA.  
 XX AC AAU96327;  
 XX DT 15-JUL-2002 (first entry)  
 XX DE Der HMW-map polypeptide #14.  
 XX DE Der HMW-map; American house dust mite; anti-allergic; mite; IgE;  
 KM mite allergenic protein; immunoglobulin E; hypersensitivity;  
 KM immunocomplex formation.  
 XX OS Dermatophagoides farinae.  
 XX PH Key Location/Qualifiers  
 FT Peptide 1..19  
 FT /note= "Signal peptide"  
 FT Protein 20..555  
 FT /note= "Mature PDerf98-555"

XX 14-SEP-2000; 2000US-00662293.  
 PR (HESK-) HESKA CORP.  
 XX Mccall CA, Hunter SW, Weber ER;  
 XX WPI; 2002-351888/38.  
 DR N-PSDB; AAK69571.  
 XX  
 PT New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HMW-map protein, useful as a vaccine for treating mite allergy.  
 XX  
 PS Claim 12; Page 114-116; 161pp; English.  
 XX  
 CC The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic  
 CC acid. The Der HMW-map protein is useful for eliciting an immune response  
 CC against Der HMW-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitizing a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig) E or Der HMW-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HMW-map polypeptides of the invention  
 CC  
 SQ Sequence 555 AA;  
 XX  
 Query Match 94.5%; Score 103; DB 5; Length 555;  
 Best Local Similarity 94.7%; Pred. No. 1.3e-08;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 DIPPTNTHKYLVCSEVNG 19  
 DB 512 DIPPTNTHKYLVCSEVNG 530  
 XX  
 RESULT 9  
 AAU96328  
 ID AAU96328 standard; protein; 555 AA.  
 XX  
 AC AAU96328;  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX  
 DE Der HMW-map polypeptide #15.  
 XX  
 KW Der HMW-map; American house dust mite; antiallergic; mite; IgE;  
 KW mite allergenic protein; immunoglobulin E; hypersensitivity;  
 KW immunocomplex formation.  
 XX  
 OS Dermatophagoides farinae.  
 XX  
 PN WO200222807-A2.  
 XX  
 PD 21-MAR-2002.  
 XX  
 PF 14-SEP-2001; 2001WO-US028730.  
 XX  
 PR 14-SEP-2000; 2000US-00662293.  
 XX  
 PA (HESK-) HESKA CORP.  
 XX  
 PI Mccall CA, Hunter SW, Weber ER;  
 XX  
 DR WPI; 2002-351888/38.  
 DR N-PSDB; AAK69573.  
 XX

PT New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HMW-map protein, useful as a vaccine for treating mite allergy.  
 XX  
 PS Claim 12; Page 120-122; 161pp; English.  
 XX  
 CC The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic  
 CC acid. The Der HMW-map protein is useful for eliciting an immune response  
 CC against Der HMW-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitizing a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig) E or Der HMW-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HMW-map polypeptides of the invention  
 CC  
 SQ Sequence 555 AA;  
 XX  
 Query Match 94.5%; Score 103; DB 5; Length 555;  
 Best Local Similarity 94.7%; Pred. No. 1.3e-08;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 DIPPTNTHKYLVCSEVNG 19  
 DB 512 DIPPTNTHKYLVCSEVNG 530  
 XX  
 RESULT 10  
 AAU96335  
 ID AAU96335 standard; protein; 490 AA.  
 XX  
 AC AAU96335;  
 XX  
 DT 06-AUG-2003 (revised)  
 DT 22-FEB-2000 (first entry)  
 XX  
 DE D. pteronyssinus 98 kD mite allergen protein (map) pDerp98-490.  
 XX  
 KW Mite allergen protein; map; high molecular weight; HMW-map; allergy;  
 KW house dust mite; IgE; immunoglobulin E; allergen; map;  
 KW hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
 KW canine; veterinary; antibody; vaccine; immunisation.  
 XX  
 OS Dermatophagoides pteronyssinus.  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 115..117  
 FT Modified-site /note="Asn is N-glycosylated"  
 FT Modified-site 240..242  
 FT /note="Asn is N-glycosylated"  
 XX  
 PN WO9954349-A2.  
 XX  
 PD 28-OCT-1999.  
 XX  
 PF 16-APR-1999; 99WO-US008524.  
 XX  
 PR 17-APR-1998; 98US-00062013.  
 PR 13-MAY-1998; 98US-0085295P.  
 PR 02-SEP-1998; 98US-0098909P.  
 XX  
 PA (HESK-) HESKA CORP.  
 XX  
 PI Mccall CA, Hunter SW, Weber ER;  
 XX  
 DR WPI; 2000-052700/04.  
 DR N-PSDB; AA236589, AA236590.  
 XX

PT Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
 PT used to modify an animals' hypersensitivity to mite allergens.

PS Claim 3; Page 147-149; 154pp; English.

XX This sequence represents Dermatophagoides pteronyssinus mite allergen  
 CC protein (map) Pderp98-490, the mature form of pderp98-509. Pderp98-490  
 CC has a molecular weight of 98 kD, comprising 490 amino acids, and has a  
 CC high degree of homology with the D. farinae mature 98 kD allergen, map8  
 CC (AA52525). Nucleic acid molecules encoding Pderp98-490 were isolated  
 CC from a D. pteronyssinus cDNA library by hybridisation with a probe  
 CC encoding the D. farinae high molecular weight map (HWM-map) composition.  
 CC Mite allergenic proteins and peptides, and nucleic acids encoding them,  
 CC may be used in therapeutic compositions to modify an animal's  
 CC hypersensitivity reaction to mite allergens. Animals that may be treated  
 CC include mammals and birds, especially felines, canines, equines, humans,  
 CC other pets, and work or domestic animals. The proteins or fragments may  
 CC also be used to diagnose allergies via a skin test. The proteins and  
 CC peptides can also be used to raise antibodies, which have a variety of  
 CC potential uses. For example, they can be used as vaccines to passively  
 CC immunise animals against dust mite hypersensitivity, as positive controls  
 CC in test kits and as tools to recover desired dust mite allergens from a  
 CC mixture of proteins. (Updated on 06-AUG-2003 to correct OS field.)

XX Sequence 490 AA;

Query Match 70.2%; Score 76.5; DB 3; Length 490;

Best Local Similarity 61.9%; Pred. No. 0.00036;  
 Matches 13; Conservative 4; Mismatches 1; Indels 3; Gaps 1;

QY 2 IPHPTNKHVLCESV---NG 19  
 :||||:|||||:|||||  
 Db 445 LPHPTDVHKLVCXYATPNG 465

RESULT 11

ID AAU96339 standard; protein; 490 AA.

XX AAU96339;

DT 15-JUL-2002 (first entry)

DE Der HWM-map polypeptide #26.

KW Der HWM-map; American house dust mite; antiallergic; mite; IgE;  
 KW mite allergenic protein; immunoglobulin E; hypersensitivity;  
 KW immunocomplex formation.

XX Dermatophagoides farinae.

OS WO200222807-A2.

PN 21-MAR-2002.

PF 14-SEP-2001; 2001WO-US028730.

PR 14-SEP-2000; 2000US-00662293.

XX (HESK-) HESKA CORP.

PA Mccall CA, Hunter SW, Weber ER;

PI WPI; 2002-351886/38.

DR N-PSDB; ABR69585.

XX New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HWM-map protein, useful as a vaccine for treating mite allergy.

PS Claim 12; Page 144-146; 161pp; English.

XX The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HWM-map protein, and its related nucleic

CC acid. The Der HWM-map protein is useful for eliciting an immune response  
 CC against Der HWM-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitising a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HWM-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig) E or Der HWM-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HWM-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HWM-map polypeptides of the invention

XX Sequence 490 AA;

Query Match 70.2%; Score 76.5; DB 5; Length 490;

Best Local Similarity 61.9%; Pred. No. 0.00036;  
 Matches 13; Conservative 4; Mismatches 1; Indels 3; Gaps 1;

QY 2 IPHPTNKHVLCESV---NG 19  
 :||||:|||||:|||||  
 Db 445 LPHPTDVHKLVCXYATPNG 465

RESULT 12

ID AA52533 standard; protein; 509 AA.

XX AA52533;

DT 06-AUG-2003 (revised)

DT 22-FEB-2000 (first entry)

DE D. pteronyssinus 98 kD mite allergen protein (map) Pderp98-509.

KW Mite allergen protein; map; high molecular weight; HWM-map; allergy;  
 KW house dust mite; IgE; immunoglobulin E; allergen; map8;  
 KW hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
 KW canine; veterinary; antibody; vaccine; immunisation.

XX Dermatophagoides pteronyssinus.

OS Key Location/Qualifiers

FT Peptide 1..19 /note="Signal peptide"

FT Protein 20..509 /note="Mature Pderp98-509"

XX WO9954349-A2.

PN 28-OCT-1999.

PF 16-APR-1999; 99WO-US008524.

PR 17-APR-1998; 98US-00062013.

PR 13-MAY-1998; 98US-0085295P.

PR 02-SEP-1998; 98US-0098909P.

XX (HESK-) HESKA CORP.

PA Mccall CA, Hunter SW, Weber ER;

PI WPI; 2000-052700/04.

DR N-PSDB; AA38585, AA38586, AA38587, AA38588.

XX Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
 PT used to modify an animals' hypersensitivity to mite allergens.

PS Claim 3; Page 134-136; 154pp; English.

XX This sequence represents Dermatophagoides pteronyssinus mite allergen  
 CC protein (map) Pderp98-509. Pderp98-509 has a molecular weight of 98 kD,

CC comprising 509 amino acids, and has a high degree of homology with the D. farinae 98 kD allergen, mapB (AF152523). Nucleic acid molecules encoding CC Pdeip98-509 were isolated from a D. pteronyssinus cDNA library by CC hybridisation with a probe encoding the D. farinae high molecular weight CC map (HMM-map) composition. Mite allergenic proteins and peptides, and CC nucleic acids encoding them, may be used in therapeutic compositions to CC modify an animal's hypersensitivity reaction to mite allergens. Animals CC that may be treated include mammals and birds, especially felines, CC canines, equines, humans, other pets, and work or domestic animals. The CC proteins or fragments may also be used to diagnose allergies via a skin CC test. The proteins and peptides can also be used to raise antibodies, CC which have a variety of potential uses. For example, they can be used as CC vaccines to passively immunise animals against dust mite CC hypersensitivity, as positive controls in test kits and as tools to CC recover desired dust mite allergens from a mixture of proteins. (Updated CC on 06-ANG-2003 to correct OS field.)

XX Sequence 509 AA;

Query Match 70.2%; Score 76.5; DB 3; Length 509;

Best Local Similarity 61.9%; Pred. No. 0.00038;

Matches 13; Conservative 4; Mismatches 1; Indels 3; Gaps 1;

QY 2 IPHPTNIHKYLVCESV---NG 19

DB 464 LPHTDVHKYLVCEYATPNG 484

RESULT 13

AAU96337

ID AAU96337 standard; protein; 509 AA.

XX AAU96337;

DT 15-JUL-2002 (first entry)

DE Der HMM-map polypeptide #24.

XX Der HMM-map; American house dust mite; antiallergic; mite; IgE;

KW mite allergenic protein; immunoglobulin E; hypersensitivity;

KW immunocomplex formation.

OS Dermatophagoides farinae.

PN WO200222807-A2.

PD 21-MAR-2002.

PF 14-SEP-2001; 2001WO-US028730.

PR 14-SEP-2000; 2000US-00662293.

PA (HESK-) HESKA CORP.

PI McCall CA, Hunter SW, Weber ER;

DR WPI; 2002-351888/38.

DR N-PSDB; ABK69581.

PT New mite allergenic protein isolated from Dermatophagoides, designated  
PT Der HMM-map protein, useful as a vaccine for treating mite allergy.

XX Claim 12; Page 134-136; 161pp; English.

CC The invention relates to an isolated mite allergenic protein of  
CC Dermatophagoides, designated Der HMM-map protein, and its related nucleic  
CC acid. The Der HMM-map protein is useful for eliciting an immune response  
CC against Der HMM-map protein. The protein or a reagent comprising a non-  
CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
CC cat) susceptible to or having an allergic response to a mite. A  
CC therapeutic composition is useful for desensitising a host animal to an  
CC allergic response to a mite. The DNA and protein can be used in the  
CC detection of anti-Der HMM-map antibodies in animal fluids, and inhibition

CC of immunoglobulin (Ig)E or Der HMM-map protein activity associated with a  
CC disease. Antibodies that bind to Der HMM-map are useful for inhibiting  
CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
CC reducing hypersensitivity responses to mite allergens, and as vaccines  
CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
CC represent Der HMM-map polypeptides of the invention

XX Sequence 509 AA;

Query Match 70.2%; Score 76.5; DB 5; Length 509;

Best Local Similarity 61.9%; Pred. No. 0.00038;

Matches 13; Conservative 4; Mismatches 1; Indels 3; Gaps 1;

QY 2 IPHPTNIHKYLVCESV---NG 19

DB 464 LPHTDVHKYLVCEYATPNG 484

RESULT 14

AAU96338

ID AAU96338 standard; protein; 509 AA.

XX AAU96338;

DT 15-JUL-2002 (first entry)

DE Der HMM-map polypeptide #25.

XX Der HMM-map; American house dust mite; antiallergic; mite; IgE;

KW mite allergenic protein; immunoglobulin E; hypersensitivity;

KW immunocomplex formation.

OS Dermatophagoides farinae.

PN WO200222807-A2.

PD 21-MAR-2002.

PF 14-SEP-2001; 2001WO-US028730.

PR 14-SEP-2000; 2000US-00662293.

PA (HESK-) HESKA CORP.

PI McCall CA, Hunter SW, Weber ER;

DR WPI; 2002-351888/38.

DR N-PSDB; ABK69583.

PT New mite allergenic protein isolated from Dermatophagoides, designated  
PT Der HMM-map protein, useful as a vaccine for treating mite allergy.

XX Claim 12; Page 139-141; 161pp; English.

CC The invention relates to an isolated mite allergenic protein of  
CC Dermatophagoides, designated Der HMM-map protein, and its related nucleic  
CC acid. The Der HMM-map protein is useful for eliciting an immune response  
CC against Der HMM-map protein. The protein or a reagent comprising a non-  
CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
CC cat) susceptible to or having an allergic response to a mite. A  
CC therapeutic composition is useful for desensitising a host animal to an  
CC allergic response to a mite. The DNA and protein can be used in the  
CC detection of anti-Der HMM-map antibodies in animal fluids, and inhibition  
CC of immunoglobulin (Ig)E or Der HMM-map protein activity associated with a  
CC disease. Antibodies that bind to Der HMM-map are useful for inhibiting  
CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
CC reducing hypersensitivity responses to mite allergens, and as vaccines  
CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
CC represent Der HMM-map polypeptides of the invention

XX Sequence 509 AA;

Query Match 70.2%; Score 76.5; DB 5; Length 509;

Best Local Similarity 61.5%; Pred. No. 0.00038;  
Matches 13; Conservative 4; Mismatches 1; Indels 3; Gaps 1;

QY 2 IPHPTNIHKYLVCESY--NG 19  
:||||:|||||:|||||  
DB 464 LPHPTDVHKYLVCEYIATPNG 464

## RESULT 15

AAO17674  
ID AAO17674 standard; protein; 143 AA.

AC AAO17674;

DT 05-AUG-2002 (first entry)

DE B tropicalis allergen variant Biot 12(S) protein SEQ ID NO: 19.

KM Mite; allergen; Biot 3; Biot 12(S); Biot 12; dust mite; immunisation;  
KM antiallergic; anti-inflammatory; immunomodulatory; storage mite; asthma;  
KM atopic dermatitis; rhinitis; gene therapy; vaccine.

OS Blomia tropicalis.

PN MO200230968-A1.

PD 18-APR-2002.

PF 12-OCT-2001; 2001WO-AU001286.

PR 13-OCT-2000; 2000AU-00000730.

PR 13-OCT-2000; 2000AU-00000731.

PA (UYST-) UNIV SINGAPORE NAT.

PI Chua KY, Nge C, Lee BW;

DR WPI: 2002-435438/46.

DR N-PSDB; AAL46687.

PT New protein allergens from the mite Blomia tropicalis, useful as vaccine  
PT for preventing, reducing or ameliorating a B. tropicalis hypersensitivity  
PT or allergic conditions, e.g. asthma, atopic dermatitis or rhinitis.

PS Claim 15; Page 133; 150pp; English.

CC The present invention relates to isolated protein allergens from the  
CC Blomia tropicalis mite, designated Biot 3 and Biot 12(S). The protein  
CC allergens are useful for preventing, reducing or ameliorating a B.  
CC tropicalis hypersensitivity condition, such as asthma, atopic dermatitis  
CC or rhinitis. The present sequence is a protein described in the invention

SQ Sequence 143 AA;

Query Match 61.5%; Score 67; DB 5; Length 143;

Best Local Similarity 69.2%; Pred. No. 0.0036;  
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 IPHPTNIHKYLVCE 14  
:||||:|||||:|||||  
DB 100 IPHPTNIHKYLVCE 112

Search completed: March 22, 2004, 06:51:43  
Job time : 6.7771 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:59:34 ; Search time 3.8866 Seconds  
(without alignments)  
1265.926 Million cell updates/sec

Title: US-09-662-293-11

Perfect score: 109  
Sequence: 1 DIPHPNTHKIVCESVNG 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1049977 seqs, 258955339 residues

Total number of hits satisfying chosen parameters: 1049977

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: Published Applications AA: \*  
2: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep: \*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep: \*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep: \*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep: \*  
6: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep: \*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep: \*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep: \*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep: \*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep: \*  
11: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep: \*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep: \*  
13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep: \*  
14: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep: \*  
15: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep: \*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep: \*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep: \*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	109	100.0	19	14	US-10-218-743-11 Sequence 11, Appl
2	109	100.0	20	14	US-10-218-743-3 Sequence 3, Appl
3	103	94.5	536	14	US-10-218-743-21 Sequence 21, Appl
4	103	94.5	555	14	US-10-218-743-15 Sequence 15, Appl
5	103	94.5	555	14	US-10-218-743-18 Sequence 18, Appl
6	76.5	70.2	490	14	US-10-218-743-41 Sequence 41, Appl
7	76.5	70.2	509	14	US-10-218-743-35 Sequence 35, Appl
8	76.5	70.2	509	14	US-10-218-743-38 Sequence 38, Appl
9	54.1	70.2	67	11	US-09-864-408A-5584 Sequence 5584, Ap
10	52	47.7	60	12	US-10-424-599-272856 Sequence 272856,
11	47	43.1	291	12	US-10-424-599-260412 Sequence 260412,
12	47	43.1	291	14	US-10-222-723-17 Sequence 17, Appl
13	46	42.2	95	12	US-10-424-599-232599 Sequence 232599,
14	46	42.2	156	12	US-10-424-599-260414 Sequence 260414,
15	46	42.2	171	12	US-10-425-114-69228 Sequence 69228, A

16	46	42.2	1344	9	US-09-738-626-6888 Sequence 6888, Ap
17	45	41.3	592	14	US-10-288-556-18 Sequence 18, Appl
18	45	41.3	867	16	US-10-389-566-1146 Sequence 1146, Ap
19	45	41.3	2327	12	US-10-016-248-55 Sequence 55, Appl
20	44.5	40.8	503	12	US-10-425-114-65522 Sequence 65522, A
21	44	40.4	410	12	US-10-424-599-274767 Sequence 274767,
22	44	40.4	713	16	US-10-354-437-64 Sequence 64, Appl
23	43.5	39.9	290	12	US-10-424-599-144566 Sequence 144566,
24	43.5	39.9	382	12	US-10-425-114-59123 Sequence 59123, A
25	43.5	39.9	398	12	US-10-425-114-59943 Sequence 59943, A
26	43.5	39.9	102	12	US-10-425-114-59943 Sequence 59943, A
27	43	39.4	232	14	US-10-282-122A-76414 Sequence 76414, A
28	43	39.4	232	15	US-10-370-570-64 Sequence 64, Appl
29	43	39.4	452	8	US-08-841-636A-35 Sequence 35, Appl
30	43	39.4	495	12	US-10-282-122A-65562 Sequence 65562, A
31	43	39.4	502	15	US-10-369-493-18401 Sequence 18401, A
32	43	39.4	897	15	US-10-369-493-18525 Sequence 18525, A
33	42.5	39.0	297	12	US-10-424-599-277795 Sequence 277795,
34	42	38.5	61	12	US-10-424-599-218599 Sequence 218599,
35	42	38.5	72	13	US-10-001-887-131 Sequence 131, App
36	42	38.5	148	12	US-10-424-599-263786 Sequence 263786,
37	42	38.5	204	11	US-09-833-245-1212 Sequence 1212, Ap
38	42	38.5	227	9	US-09-925-300-1124 Sequence 1124, Ap
39	42	38.5	270	12	US-10-425-114-38713 Sequence 38713, A
40	42	38.5	289	11	US-09-833-245-1214 Sequence 1214, Ap
41	42	38.5	291	9	US-10-222-723-10 Sequence 10, Appl
42	42	38.5	319	9	US-09-971-361-9 Sequence 9, Appl
43	42	38.5	319	9	US-09-971-361-9 Sequence 9, Appl
44	42	38.5	451	12	US-10-424-599-144134 Sequence 144134,
45	42	38.5	506	12	US-10-424-599-144132 Sequence 144132,

#### ALIGNMENTS

RESULT 1  
US-10-218-743-11  
Sequence 11, Application US/10218743  
Publication No. US20030096779A1  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley W.  
APPLICANT: Weber, Eric R.  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
TITLE OR INVENTION: AND USES THEREOF  
FILE REFERENCE: AU-2-C3  
CURRENT FILING DATE: 2002-08-13  
PRIOR APPLICATION NUMBER: US/09/232,225  
PRIOR FILING DATE: 1999-04-15  
PRIOR APPLICATION NUMBER: 60/098,909  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/085,295  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/098,565  
PRIOR FILING DATE: 1998-04-17  
PRIOR APPLICATION NUMBER: 09/062,013  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 11  
LENGTH: 19  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-10-218-743-11  
Query Match 100.0%; Score 109; DB 14; Length 19;  
Best Local Similarity 100.0%; Pred. No. 7; 5e-10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cy 1 DIPHPNTHKIVCESVNG 19  
Db 1 DIPHPNTHKIVCESVNG 19

```
RESULT 2
US-10-218-743-3
; Sequence 3, Application US/10218743
; Publication No. US20030096779A1
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/10/218,743
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US/09/292,225
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: 60/098,909
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/085,295
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/098,565
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: 09/062,013
; PRIOR FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-10-218-743-3

Query Match          100.0%; Score 109; DB 14; Length 20;
Best Local Similarity 100.0%; Pred. No. 7,9e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DHPPTNHHKYLVCESVNG 19
DB      1 DHPPTNHHKYLVCESVNG 19

RESULT 3
US-10-218-743-21
; Sequence 21, Application US/10218743
; Publication No. US20030096779A1
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/10/218,743
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US/09/292,225
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: 60/098,909
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/085,295
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/098,565
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: 09/062,013
; PRIOR FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-10-218-743-21
```

```
Query Match          94.5%; Score 103; DB 14; Length 536;
Best Local Similarity 94.7%; Pred. No. 2,1e-07;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 DHPPTNHHKYLVCESVNG 19
DB      493 DHPPTNHHKYLVCESVNG 511

RESULT 4
US-10-218-743-15
; Sequence 15, Application US/10218743
; Publication No. US20030096779A1
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/10/218,743
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US/09/292,225
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: 60/098,909
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/085,295
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/098,565
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: 09/062,013
; PRIOR FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-10-218-743-15

Query Match          94.5%; Score 103; DB 14; Length 555;
Best Local Similarity 94.7%; Pred. No. 2,1e-07;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 DHPPTNHHKYLVCESVNG 19
DB      512 DHPPTNHHKYLVCESVNG 530

RESULT 5
US-10-218-743-18
; Sequence 18, Application US/10218743
; Publication No. US20030096779A1
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/10/218,743
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: US/09/292,225
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: 60/098,909
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/085,295
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/098,565
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: 09/062,013
; PRIOR FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
```



SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 18  
LENGTH: 555  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-10-218-743-18

Query Match 94.5%; Score 103; DB 14; Length 555;  
Best Local Similarity 94.7%; Pred. No. 2.1e-07;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DHPPTNIHKYLVCEVNG 19  
|||||  
Db 512 DHPPTNIHKYLVCEVNG 530

RESULT 6  
US-10-218-743-41  
Sequence 41, Application US/10218743  
Publication No. US20030096779A1  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley Wu  
APPLICANT: Weber, Eric R.  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/10/218,743  
CURRENT FILING DATE: 2002-08-13  
PRIOR APPLICATION NUMBER: US/09/292,225  
PRIOR FILING DATE: 1999-04-15  
PRIOR APPLICATION NUMBER: 60/098,909  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/085,295  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/098,565  
PRIOR FILING DATE: 1998-04-17  
PRIOR APPLICATION NUMBER: 09/062,013  
PRIOR FILING DATE: 1998-04-17  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 41  
LENGTH: 490  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-10-218-743-41

Query Match 70.2%; Score 76.5; DB 14; Length 490;  
Best Local Similarity 61.9%; Pred. No. 0.0022;  
Matches 13; Conservative 4; Mismatches 1; Indels 3; Gaps 1;

QY 2 IPHPTNIHKYLVCESV--NG 19  
:|||||:|||||:  
Db 445 LPHPTDVHKYLVCEYIATPNG 465

RESULT 7  
US-10-218-743-35  
Sequence 35, Application US/10218743  
Publication No. US20030096779A1  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley Wu  
APPLICANT: Weber, Eric R.  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/10/218,743  
CURRENT FILING DATE: 2002-08-13  
PRIOR APPLICATION NUMBER: US/09/292,225  
PRIOR FILING DATE: 1999-04-15  
PRIOR APPLICATION NUMBER: 60/098,909  
PRIOR FILING DATE: 1998-09-02

PRIOR APPLICATION NUMBER: 60/085,295  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/098,565  
PRIOR FILING DATE: 1998-04-17  
PRIOR APPLICATION NUMBER: 09/062,013  
PRIOR FILING DATE: 1998-04-17  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 35  
LENGTH: 509  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-10-218-743-35

Query Match 70.2%; Score 76.5; DB 14; Length 509;  
Best Local Similarity 61.9%; Pred. No. 0.0023;  
Matches 13; Conservative 4; Mismatches 1; Indels 3; Gaps 1;

QY 2 IPHPTNIHKYLVCESV--NG 19  
:|||||:|||||:  
Db 464 LPHPTDVHKYLVCEYIATPNG 484

RESULT 8  
US-10-218-743-38  
Sequence 38, Application US/10218743  
Publication No. US20030096779A1  
GENERAL INFORMATION:  
APPLICANT: McCall, Catherine A.  
APPLICANT: Hunter, Shirley Wu  
APPLICANT: Weber, Eric R.  
TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
FILE REFERENCE: AL-2-C3  
CURRENT APPLICATION NUMBER: US/10/218,743  
CURRENT FILING DATE: 2002-08-13  
PRIOR APPLICATION NUMBER: US/09/292,225  
PRIOR FILING DATE: 1999-04-15  
PRIOR APPLICATION NUMBER: 60/098,909  
PRIOR FILING DATE: 1998-09-02  
PRIOR APPLICATION NUMBER: 60/085,295  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/098,565  
PRIOR FILING DATE: 1998-04-17  
PRIOR APPLICATION NUMBER: 09/062,013  
PRIOR FILING DATE: 1998-04-17  
NUMBER OF SEQ ID NOS: 49  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 38  
LENGTH: 509  
TYPE: PRT  
ORGANISM: Dermatophagoides farinae  
US-10-218-743-38

Query Match 70.2%; Score 76.5; DB 14; Length 509;  
Best Local Similarity 61.9%; Pred. No. 0.0023;  
Matches 13; Conservative 4; Mismatches 1; Indels 3; Gaps 1;

QY 2 IPHPTNIHKYLVCESV--NG 19  
:|||||:|||||:  
Db 464 LPHPTDVHKYLVCEYIATPNG 484

RESULT 9  
US-09-664-408A-5584  
Sequence 5584, Application US/09864408A  
Publication No. US20040009474A1  
GENERAL INFORMATION:  
APPLICANT: Leach, Martin D.  
APPLICANT: Shinkets, Richard A.  
TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Enc  
FILE REFERENCE: 21402-012  
CURRENT APPLICATION NUMBER: US/09/864,408A

CURRENT FILING DATE: 2001-05-24  
PRIOR APPLICATION NUMBER: 60/206,690  
PRIOR FILING DATE: 2000-05-24  
NUMBER OF SEQ ID NOS: 9068  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5584  
LENGTH: 67  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-864-408A-5584

Query Match 54.1%; Score 59; DB 11; Length 67;  
Best Local Similarity 56.2%; Pred. No. 0.13;  
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 IHPHNIHKYLVCEV 17  
DB 7 IPPPANHGHVCEST 22

RESULT 10  
US-10-424-599-272856  
Sequence 272856, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 272856  
LENGTH: 60  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_88411C.1.pep  
US-10-424-599-272856

Query Match 47.7%; Score 52; DB 12; Length 60;  
Best Local Similarity 77.8%; Pred. No. 1.4;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 DIPHPNINH 9  
DB 38 DVHPNINH 46

RESULT 11  
US-10-424-599-260412  
Sequence 260412, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 260412  
LENGTH: 291  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_77175C.1.pep

US-10-424-599-260412

Query Match 43.1%; Score 47; DB 12; Length 291;  
Best Local Similarity 47.1%; Pred. No. 44;  
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 DIPHPNINHKLVCESV 17  
DB 88 DIPVTVHESIQCBAV 104

RESULT 12  
US-10-222-723-17  
Sequence 17, Application US/10222723  
Publication No. US20030145351A1  
GENERAL INFORMATION:  
APPLICANT: Stephen M. Allen  
APPLICANT: Saverio C. Falco  
APPLICANT: Dennis Flint  
APPLICANT: Steven Gutierrez  
TITLE OF INVENTION: 3-Deoxy-D-Manno-Octulosonic Acid 8-Phosphate Synthases  
FILE REFERENCE: B81512 USNA  
CURRENT APPLICATION NUMBER: US/10/222,723  
CURRENT FILING DATE: 2002-08-16  
PRIOR APPLICATION NUMBER: 60/312,679  
PRIOR FILING DATE: 2001-08-16  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: Microsoft Word Version 7.0A  
SEQ ID NO 17  
LENGTH: 291  
TYPE: PRT  
ORGANISM: Glycine max  
US-10-222-723-17

Query Match 43.1%; Score 47; DB 14; Length 291;  
Best Local Similarity 47.1%; Pred. No. 44;  
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 DIPHPNINHKLVCESV 17  
DB 88 DIPVTVHESIQCBAV 104

RESULT 13  
US-10-424-599-232599  
Sequence 232599, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 232599  
LENGTH: 95  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_52061C.1.pep  
US-10-424-599-232599

Query Match 42.2%; Score 46; DB 12; Length 95;  
Best Local Similarity 50.0%; Pred. No. 19;  
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 4 HPTNHHKYLVCESV 17  
DB 33 HITHHKEVACESL 46

RESULT 14  
US-10-424-599-260414  
; Sequence 260414, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 260414  
; LENGTH: 156  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_77177C.1.pap  
US-10-424-599-260414

Query Match 42.2%; Score 46; DB 12; Length 156;  
Best Local Similarity 47.1%; Pred. No. 32;  
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 DIPHTNIHKYLVCEV 17  
||| |:::| |::|  
Db 88 DIPITDVHETHCEAV 104

RESULT 15  
US-10-425-114-69228  
; Sequence 69228, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 69228  
; LENGTH: 171  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17103F03\_FLI.pap  
US-10-425-114-69228

Query Match 42.2%; Score 46; DB 12; Length 171;  
Best Local Similarity 58.3%; Pred. No. 36;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 PHTNIHKYLVLC 14  
||| |:::| |::|  
Db 33 PHTTHHYSLC 44

Search completed: March 22, 2004, 07:45:49  
Job time : 3.8866 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:42:54 ; Search time 1.53141 Seconds  
(without alignments)  
640.518 Million cell updates/sec

Title: US-09-662-293-11  
Perfect score: 103  
Sequence: 1 DIPHTNIHKYLCESVNG 19

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PTCUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	109	100.0	19	4	US-09-292-225-11
2	109	100.0	20	4	US-09-292-225-3
3	103	94.5	536	4	US-09-292-225-21
4	103	94.5	555	4	US-09-292-225-15
5	103	94.5	555	4	US-09-292-225-18
6	76.5	70.2	490	4	US-09-292-225-41
7	76.5	70.2	509	4	US-09-292-225-35
8	76.5	70.2	509	4	US-09-292-225-38
9	55	50.5	517	4	US-09-257-825B-21
10	46	42.2	520	4	US-09-257-825B-20
11	45	41.3	592	4	US-08-533-711B-18
12	43.5	39.9	232	2	US-08-738-413B-11
13	43	39.4	259	4	US-09-328-352-5038
14	43	39.4	452	3	US-09-329-350-35
15	42	38.5	517	4	US-09-402-515A-16
16	42	38.5	374	4	US-09-489-039A-7383
17	42	38.5	1088	3	US-08-633-768A-1
18	42	38.5	1088	4	US-09-280-197-1
19	42	38.5	1092	4	US-09-275-608-3
20	42	38.5	1843	3	US-09-413-814-50
21	41.5	38.1	159	4	US-09-621-377B-4
22	41	37.6	76	6	5459061-2
23	41	37.6	103	4	US-09-732-210-1269
24	41	37.6	182	4	US-08-858-207A-434
25	41	37.6	202	1	US-08-155-171B-21
26	41	37.6	202	2	US-08-435-998-21
27	41	37.6	509	1	US-10-095-946-8

#### ALIGNMENTS

##### RESULT 1

US-09-292-225-11  
; Sequence 11, Application US/09292225  
; Patent No. 6455686  
; GENERAL INFORMATION:  
; APPLICANT: McCall, Catherine A.  
; APPLICANT: Hunter, Shirley Wu  
; APPLICANT: Weber, Eric R.  
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: AL-2-C3  
; CURRENT APPLICATION NUMBER: US/09/292,225  
; CURRENT FILING DATE: 1999-04-15  
; EARLIER APPLICATION NUMBER: 60/098,909  
; EARLIER FILING DATE: 1998-09-02  
; EARLIER APPLICATION NUMBER: 60/085,295  
; EARLIER FILING DATE: 1998-05-13  
; EARLIER APPLICATION NUMBER: 60/098,565  
; EARLIER FILING DATE: 1998-04-17  
; EARLIER APPLICATION NUMBER: 09/062,013  
; EARLIER FILING DATE: 1998-04-17  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: Patentin ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Dermatophagoides farinae  
US-09-292-225-11

Query Match 100.0%; Score 109; DB 4; Length 19;  
Best Local Similarity 100.0%; Pred. No. 1.7e-11;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DIPHTNIHKYLCESVNG 19  
|||  
Db 1 DIPHTNIHKYLCESVNG 19

##### RESULT 2

US-09-292-225-3  
; Sequence 3, Application US/09292225  
; Patent No. 6455686  
; GENERAL INFORMATION:  
; APPLICANT: McCall, Catherine A.  
; APPLICANT: Hunter, Shirley Wu  
; APPLICANT: Weber, Eric R.  
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: AL-2-C3  
; CURRENT APPLICATION NUMBER: US/09/292,225

```
/ CURRENT FILING DATE: 1999-04-15
/ EARLIER APPLICATION NUMBER: 60/098,909
/ EARLIER FILING DATE: 1998-09-02
/ EARLIER APPLICATION NUMBER: 60/085,295
/ EARLIER FILING DATE: 1998-05-13
/ EARLIER APPLICATION NUMBER: 60/098,565
/ EARLIER FILING DATE: 1998-04-17
/ EARLIER APPLICATION NUMBER: 60/062,013
/ EARLIER FILING DATE: 1998-04-17
/ NUMBER OF SEQ ID NOS: 49
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 3
/ LENGTH: 20
/ TYPE: PRT
/ ORGANISM: Dermatophagoides farinae
US-09-292-225-3

Query Match      100.0%; Score 109; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.8e-11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DIPHTNIHKYLVCSVNG 19
   |||||
Db 1 DIPHTNIHKYLVCSVNG 19
   |||||

RESULT 3
US-09-292-225-21
/ Sequence 21, Application US/09292225
/ Patent No. 6455686
/ GENERAL INFORMATION:
/ APPLICANT: McCall, Catherine A.
/ APPLICANT: Hunter, Shirley Wu
/ APPLICANT: Weber, Eric R.
/ TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
/ FILE OF INVENTION: AND USES THEREOF
/ FILE REFERENCE: AL-2-C3
/ CURRENT APPLICATION NUMBER: US/09/292,225
/ CURRENT FILING DATE: 1999-04-15
/ EARLIER APPLICATION NUMBER: 60/098,909
/ EARLIER FILING DATE: 1998-09-02
/ EARLIER APPLICATION NUMBER: 60/085,295
/ EARLIER FILING DATE: 1998-05-13
/ EARLIER APPLICATION NUMBER: 60/098,565
/ EARLIER FILING DATE: 1998-04-17
/ EARLIER APPLICATION NUMBER: 60/062,013
/ EARLIER FILING DATE: 1998-04-17
/ NUMBER OF SEQ ID NOS: 49
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 21
/ LENGTH: 536
/ TYPE: PRT
/ ORGANISM: Dermatophagoides farinae
US-09-292-225-21

Query Match      94.5%; Score 103; DB 4; Length 536;
Best Local Similarity 94.7%; Pred. No. 6.2e-09;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DIPHTNIHKYLVCSVNG 19
   |||||
Db 493 DIPHTNIHKYLVCFVNG 511
   |||||

RESULT 4
US-09-292-225-15
/ Sequence 15, Application US/09292225
/ Patent No. 6455686
/ GENERAL INFORMATION:
/ APPLICANT: McCall, Catherine A.
/ APPLICANT: Hunter, Shirley Wu
/ APPLICANT: Weber, Eric R.
/ TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
```

```
/ TITLE OF INVENTION: AND USES THEREOF
/ FILE REFERENCE: AL-2-C3
/ CURRENT APPLICATION NUMBER: US/09/292,225
/ CURRENT FILING DATE: 1999-04-15
/ EARLIER APPLICATION NUMBER: 60/098,909
/ EARLIER FILING DATE: 1998-09-02
/ EARLIER APPLICATION NUMBER: 60/085,295
/ EARLIER FILING DATE: 1998-05-13
/ EARLIER APPLICATION NUMBER: 60/098,565
/ EARLIER FILING DATE: 1998-04-17
/ EARLIER APPLICATION NUMBER: 09/062,013
/ EARLIER FILING DATE: 1998-04-17
/ NUMBER OF SEQ ID NOS: 49
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 15
/ LENGTH: 555
/ TYPE: PRT
/ ORGANISM: Dermatophagoides farinae
US-09-292-225-15

Query Match      94.5%; Score 103; DB 4; Length 555;
Best Local Similarity 94.7%; Pred. No. 6.6e-09;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DIPHTNIHKYLVCSVNG 19
   |||||
Db 512 DIPHTNIHKYLVCFVNG 530
   |||||

RESULT 5
US-09-292-225-18
/ Sequence 18, Application US/09292225
/ Patent No. 6455686
/ GENERAL INFORMATION:
/ APPLICANT: McCall, Catherine A.
/ APPLICANT: Hunter, Shirley Wu
/ APPLICANT: Weber, Eric R.
/ TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
/ FILE OF INVENTION: AND USES THEREOF
/ FILE REFERENCE: AL-2-C3
/ CURRENT APPLICATION NUMBER: US/09/292,225
/ CURRENT FILING DATE: 1999-04-15
/ EARLIER APPLICATION NUMBER: 60/098,909
/ EARLIER FILING DATE: 1998-09-02
/ EARLIER APPLICATION NUMBER: 60/085,295
/ EARLIER FILING DATE: 1998-05-13
/ EARLIER APPLICATION NUMBER: 60/098,565
/ EARLIER FILING DATE: 1998-04-17
/ EARLIER APPLICATION NUMBER: 09/062,013
/ EARLIER FILING DATE: 1998-04-17
/ NUMBER OF SEQ ID NOS: 49
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 18
/ LENGTH: 555
/ TYPE: PRT
/ ORGANISM: Dermatophagoides farinae
US-09-292-225-18

Query Match      94.5%; Score 103; DB 4; Length 555;
Best Local Similarity 94.7%; Pred. No. 6.6e-09;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DIPHTNIHKYLVCSVNG 19
   |||||
Db 512 DIPHTNIHKYLVCFVNG 530
   |||||

RESULT 6
US-09-292-225-41
/ Sequence 41, Application US/09292225
/ Patent No. 6455686
/ GENERAL INFORMATION:
/ APPLICANT: McCall, Catherine A.
```

```

; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; CURRENT FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: 60/098,909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085,295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098,565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062,013
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41
; LENGTH: 490
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-09-292-225-41

Query Match      70.2%; Score 76.5; DB 4; Length 490;
Best Local Similarity 61.9%; Pred. No. 0.00014;
Matches 13; Conservative 4; Mismatches 1; Indels 3; Gaps 1;

QY 2 IPHPTNIHKYLVCSV---NG 19
:||||:|||||:
Db 445 LPHTDVHKYLVCEYIATPENG 465

RESULT 7
US-09-292-225-35
; Sequence 35, Application US/09292225
; Patent No. 6455686
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; CURRENT FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: 60/098,909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085,295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098,565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062,013
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-09-292-225-35

Query Match      70.2%; Score 76.5; DB 4; Length 509;
Best Local Similarity 61.9%; Pred. No. 0.00014;
Matches 13; Conservative 4; Mismatches 1; Indels 3; Gaps 1;

QY 2 IPHPTNIHKYLVCSV---NG 19
:||||:|||||:
Db 464 LPHTDVHKYLVCEYIATPENG 484

RESULT 8
US-09-292-225-38
; Sequence 38, Application US/09292225
; Patent No. 6455686
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; CURRENT FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: 60/098,909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085,295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098,565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062,013
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-09-292-225-38

Query Match      70.2%; Score 76.5; DB 4; Length 509;
Best Local Similarity 61.9%; Pred. No. 0.00014;
Matches 13; Conservative 4; Mismatches 1; Indels 3; Gaps 1;

QY 2 IPHPTNIHKYLVCSV---NG 19
:||||:|||||:
Db 464 LPHTDVHKYLVCEYIATPENG 484

RESULT 9
US-09-257-825B-21
; Sequence 21, Application US/09257825B
; Patent No. 6403352
; GENERAL INFORMATION:
; APPLICANT: Poovaiah, Bachettira W.
; APPLICANT: Patil, Shameekumar
; APPLICANT: Takezawa, Daisuke
; TITLE OF INVENTION: Compositions and Methods for Production of Male-Sterile Plants
; FILE REFERENCE: 4630-51993
; CURRENT APPLICATION NUMBER: US/09/257,825B
; CURRENT FILING DATE: 1999-02-25
; PRIOR APPLICATION NUMBER: US 08/655,352
; PRIOR FILING DATE: 1996-05-23
; PRIOR APPLICATION NUMBER: US 60/014,743
; PRIOR FILING DATE: 1996-03-28
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 517
; TYPE: PRT
; ORGANISM: Nicotiana tabacum
US-09-257-825B-21

Query Match      50.5%; Score 55; DB 4; Length 517;
Best Local Similarity 58.8%; Pred. No. 0.51;
Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 3 PHPTNIHKYLVCSVNG 19
:||||:|||||:
Db 97 PHPNVHLVDVCEDPFG 113

RESULT 10
US-09-257-825B-20
; Sequence 20, Application US/09257825B
; Patent No. 6403352
; GENERAL INFORMATION:

```

```

; Patent No. 6455686
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; CURRENT FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: 60/098,909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085,295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098,565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062,013
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Dermatophagoides farinae
US-09-292-225-38

Query Match      70.2%; Score 76.5; DB 4; Length 509;
Best Local Similarity 61.9%; Pred. No. 0.00014;
Matches 13; Conservative 4; Mismatches 1; Indels 3; Gaps 1;

QY 2 IPHPTNIHKYLVCSV---NG 19
:||||:|||||:
Db 464 LPHTDVHKYLVCEYIATPENG 484

RESULT 9
US-09-257-825B-21
; Sequence 21, Application US/09257825B
; Patent No. 6403352
; GENERAL INFORMATION:
; APPLICANT: Poovaiah, Bachettira W.
; APPLICANT: Patil, Shameekumar
; APPLICANT: Takezawa, Daisuke
; TITLE OF INVENTION: Compositions and Methods for Production of Male-Sterile Plants
; FILE REFERENCE: 4630-51993
; CURRENT APPLICATION NUMBER: US/09/257,825B
; CURRENT FILING DATE: 1999-02-25
; PRIOR APPLICATION NUMBER: US 08/655,352
; PRIOR FILING DATE: 1996-05-23
; PRIOR APPLICATION NUMBER: US 60/014,743
; PRIOR FILING DATE: 1996-03-28
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 517
; TYPE: PRT
; ORGANISM: Nicotiana tabacum
US-09-257-825B-21

Query Match      50.5%; Score 55; DB 4; Length 517;
Best Local Similarity 58.8%; Pred. No. 0.51;
Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 3 PHPTNIHKYLVCSVNG 19
:||||:|||||:
Db 97 PHPNVHLVDVCEDPFG 113

RESULT 10
US-09-257-825B-20
; Sequence 20, Application US/09257825B
; Patent No. 6403352
; GENERAL INFORMATION:

```

APPLICANT: Poovaiyah, Bachettira W.  
APPLICANT: Patil, Shameekumar  
APPLICANT: Takezawa, Daisuke  
TITLE OF INVENTION: Compositions and Methods for Production of Male-Sterile Plants  
FILE REFERENCE: 4630-51993  
CURRENT FILING DATE: 1999-02-25  
CURRENT APPLICATION NUMBER: US/09/257,825B  
PRIOR FILING DATE: 1996-05-23  
PRIOR APPLICATION NUMBER: US 08/655,352  
PRIOR FILING DATE: 1996-03-28  
PRIOR APPLICATION NUMBER: US 60/014,743  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 20  
LENGTH: 520  
TYPE: PRT  
ORGANISM: Lilium longiflorum  
US-09-257-825B-20

Query Match 42.2%; Score 46; DB 4; Length 520;  
Best Local Similarity 52.9%; Pred. No. 16;  
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 3 PHTNHHKYLVCESVNG 19  
||| ||| : ||| |||  
Db 100 PHPNVHLDVYEDANG 116

RESULT 11  
US-08-933-711B-18  
Sequence 18, Application US/08933711B  
Patent No. 6514724  
GENERAL INFORMATION:  
APPLICANT: McMahon, Andrew P.  
APPLICANT: Chuang, Pao-Tien  
TITLE OF INVENTION: HEDGHOG INTERACTING PROTEINS AND USES RELATED THERETO  
FILE REFERENCE: HUV-024.01  
CURRENT APPLICATION NUMBER: US/08/933,711B  
CURRENT FILING DATE: 1997-09-19  
PRIOR FILING DATE: 1996-09-20  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: Patent in Ver. 2.1  
SEQ ID NO 18  
LENGTH: 592  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Consensus  
US-08-933-711B-18

Query Match 41.3%; Score 45; DB 4; Length 592;  
Best Local Similarity 38.9%; Pred. No. 27;  
Matches 7; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 IPHTNHHKYLVCESVNG 19  
: ||| ||| : ||| |||  
Db 356 VDHDPNINLTICDSNG 373

RESULT 12  
US-08-738-413B-11  
Sequence 11, Application US/08738413B  
Patent No. 5821106  
GENERAL INFORMATION:  
APPLICANT: CHUNG, Kwang-Hoe  
APPLICANT: KOH, You-Seok  
APPLICANT: HWANG, Jae-Hoon  
APPLICANT: KIM, Doo-Sik  
APPLICANT: YUN, Yung-Dae  
APPLICANT: MOON, Hong-Mo  
TITLE OF INVENTION: A NOVEL CDNA OF DIRECT-ACTING

TITLE OF INVENTION: FIBRINOLYTIC SERINE PROTEASE  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Darby & Darby PC  
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: US  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
SOFTWARE:  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/738,413B  
FILING DATE: October 23, 1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Ludwig, S. Peter  
REGISTRATION NUMBER: 25,351  
REFERENCE/DOCKET NUMBER: 0136/0C539  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-527-7700  
TELEFAX: 212-753-6237  
TELEX: 236687  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 232 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Bothrops atrox moojeni  
IMMEDIATE SOURCE:  
CLONE: BATROXOBIN  
US-08-738-413B-11

Query Match 39.9%; Score 43.5; DB 2; Length 232;  
Best Local Similarity 45.5%; Pred. No. 17;  
Matches 10; Conservative 5; Mismatches 4; Indels 3; Gaps 2;

QY 1 DIPHTNHHKYL--VC-BSVNG 19  
|: ||| ||| : ||| |||  
Db 135 DVPHCANINLFMNTVCREAYNG 156

RESULT 13  
US-09-328-352-5038  
Sequence 5038, Application US/09328352  
Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 5038  
LENGTH: 259  
TYPE: PRT  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-5038

Query Match 39.4%; Score 43; DB 4; Length 259;  
Best Local Similarity 60.0%; Pred. No. 23;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 HPTNHHKYLIV 13  
||| : ||| |||

Db 42 HPAKVHRYLV 51

RESULT 14

US-09-329-350-35

Sequence 35, Application US/09329350

Patent No. 6184019

GENERAL INFORMATION:

APPLICANT: Miettinen-Oinonen, Arja

APPLICANT: Londenborough, John

APPLICANT: Vehmaanper, Jari

APPLICANT: Haakana, Heli

APPLICANT: M ntyl , Arja

APPLICANT: Lantto, Raija

APPLICANT: Elovainio, Minna

APPLICANT: Joutsenki, Vesa

APPLICANT: Paloheimo, Marja

APPLICANT: Suominen, Pirkko

TITLE OF INVENTION: NOVEL CELLULASES, THE GENES ENCODING THEM AND

TITLE OF INVENTION: USES THEREOF

NUMBER OF SEQUENCES: 45

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.

STREET: 1100 New York Avenue, N.W., Suite 600

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/329,350

FILING DATE: Herewith

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/841,636

FILING DATE: 30-APR-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/005,335

FILING DATE: 17-OCT-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/007,926

FILING DATE: 04-DEC-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/020,840

FILING DATE: 28-JUN-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/732,181

FILING DATE: 16-OCT-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/FI96/00550

FILING DATE: 17-OCT-1996

ATTORNEY/AGENT INFORMATION:

NAME: Shea Jr., Timothy

REGISTRATION NUMBER: 41,306

REFERENCE/DOCKET NUMBER: 1716.0510006/MAC/TJS

TELEPHONE: (202)371-2600

TELEFAX: (202)371-2540

INFORMATION FOR SEQ ID NO: 35:

SEQUENCE CHARACTERISTICS:

LENGTH: 452 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: protein

ORIGINAL SOURCE:

ORGANISM: Melanocarpus albomyces

STRAIN: ALKO4237

FEATURE:

NAME/KEY: Protein

LOCATION: 1..452

OTHER INFORMATION: /label= 50K-cellulase-B

US-09-329-350-35

Query Match 39.4%; Score 43; DB 3; Length 452;

Best Local Similarity 43.8%; Pred. No. 42;

Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 3 PHPTNIHKYLVCRSVN 18

Db 249 PHACTTNEYHVCETN 264

RESULT 15

US-09-402-515A-16

Sequence 16, Application US/09402515A

Patent No. 6423316

GENERAL INFORMATION:

APPLICANT: RIESBECK, Kristian

APPLICANT: DORLING, Anthony

APPLICANT: GEORGE, Andrew

APPLICANT: LECHLER, Robert

TITLE OF INVENTION: ANTICOAGULANT FUSION PROTEIN ANCHORED TO CELL MEMBRANE

FILE REFERENCE: 2292/OG135

CURRENT APPLICATION NUMBER: US/09/402,515A

CURRENT FILING DATE: 2000-02-02

NUMBER OF SEQ ID NOS: 17

SOFTWARE: Patent In version 3.1

SEQ ID NO 16

LENGTH: 231

TYPE: PRT

ORGANISM: Agkistrodon contortrix contortrix

US-09-402-515A-16

Query Match 38.5%; Score 42; DB 4; Length 231;

Best Local Similarity 44.4%; Pred. No. 30;

Matches 8; Conservative 4; Mismatches 4; Indels 2; Gaps 1;

Qy 1 DIPHPTNIH--KYLVCES 16

Db 134 DVPHCANINILDYAVCOA 151

Search completed: March 22, 2004, 07:03:57

Job time : 1.53141 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:40:28 ; Search time 1.30072 Seconds  
(without alignments)  
1479.047 Million cell updates/sec

Title: US-09-662-293-13  
Perfect score: 105  
Sequence: 1 DPAKGMSPPGFIVGEGVLS 20  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR /8: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	50	47.6	408	2 T13387	hypothetical prote
2	47	44.8	573	2 A86287	hypothetical prote
3	46	43.8	433	2 C87552	aspartate transcar
4	45	42.9	141	2 S09804	hypothetical prote
5	44	41.9	111	2 H72501	hypothetical prote
6	44	41.9	408	2 A92338	conserved hypother
7	44	41.9	410	2 H72502	hypothetical prote
8	44	41.9	690	2 F70628	probable pta prote
9	44	41.9	813	2 T02672	hypothetical prote
10	44	41.9	835	2 I55603	reduced hepatic gl
11	44	41.9	1539	2 S65775	homeotic protein H
12	44	41.9	3224	1 S58884	Ran-binding protein
13	43	41.0	291	2 F81444	probable integral
14	43	41.0	305	2 AG2399	hypothetical prote
15	43	41.0	359	2 H70876	hypothetical prote
16	43	41.0	367	2 JC6087	helix-loop-helix t
17	43	41.0	394	2 E64028	hypothetical prote
18	43	41.0	663	2 AE3290	propionate-CoA lig
19	43	41.0	728	2 T51071	related to trfA pr
20	43	41.0	817	2 S53919	hypothetical prote
21	43	41.0	883	2 B95237	alcohol dehydrogen
22	43	41.0	890	2 C98101	alcohol-acetaldehy
23	43	41.0	891	2 AE0358	Clp ATPase [import
24	42.5	40.5	289	2 S44863	R0503.10 protein -
25	42.5	40.5	561	2 S72618	hypX protein - Rhi
26	42.5	40.5	864	2 F98196	probable clpA/B-cy
27	42.5	40.5	867	2 AH0437	Clp ATPase [import
28	42.5	40.5	869	2 G82499	clpB protein VCA01
29	42.5	40.5	892	2 AD3090	ATP-dependent Clp

30 42.5 40.5 923 2 G90556  
31 42.5 40.5 923 2 G85507  
32 42.5 40.5 3176 2 CGH3A  
33 42 40.0 190 2 A87263  
34 42 40.0 231 2 AE2164  
35 42 40.0 283 2 H84378  
36 42 40.0 438 1 WMAD40  
37 42 40.0 623 2 A49112  
38 42 40.0 677 2 I49045  
39 42 40.0 885 2 A84373  
40 42 40.0 889 2 JC5576  
41 42 40.0 907 1 A57429  
42 41.5 39.5 80 2 AF2125  
43 41.5 39.5 224 2 G87377  
44 41 39.0 184 2 AD0314  
45 41 39.0 198 2 AC3305

hypothetical prote  
probable proteinas  
collagen alpha 3(V  
hypothetical prote  
hypothetical prote  
hypothetical prote  
late 100K protein  
sodium-glucose cot  
probable ATPase SX  
leucine-tRNA synth  
inter-alpha-trypsi  
aldehyde oxidase (  
hypothetical prote  
hypothetical prote  
probable membrane  
hypothetical prote

ALIGNMENTS

RESULT 1

T13387  
hypothetical protein I15C2.8 - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 19-Jan-2001  
C:Accession: T13387  
R:Salles, C.; Valenti, P.; Darlamitsou, A.; Henderson, N.; Campbell, L.; Glover, D.  
submitted to the EMBL Data Library, May 1999  
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.  
A:Reference number: Z17665  
A:Accession: T13387  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-408 <CAT>  
A:Cross-references: EMBL:AL031581; NID:el320978; PTD:el320993; PIDN:CAA20887.1  
C:Genetics:  
A:Cross-references: FlyBase:FBgn0020381  
A:Introns: 340/1  
A:Note: EG:115C2.8  
C:Superfamily: Drosophila melanogaster hypothetical protein I15C2.8

Query Match 47.6% Score 50; DB 2; Length 408;  
Best Local Similarity 40.0%; Pred. No. 4.7;  
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Qy 1 DPAKGMSPPGFIVGEGVLS 20  
Db 327 DPQEGFDPPTLYLGDEELS 346

RESULT 2

A86287  
hypothetical protein F911.22 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001  
C:Accession: A86287  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Huijzer, L.  
ansen, N.F.; Hughes, B.; Huijzer, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: A86287  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-573 <STO>  
A:Cross-references: GB:AE005172; NID:G5103826; PIDN:AND39656.1; GSPDB:GN00141

C;Genetics:  
A;Map position: 1

Query Match 44.8%; Score 47; DB 2; Length 573;  
Best Local Similarity 47.1%; Pred. No. 21;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 PAKGMSPPGFIVGEEGV 18  
DB 477 PTFGMAYPGVQSGEDGV 493

RESULT 3  
C87552  
aspartate transcarbamoylase, pyrc' subunit [imported] - Caulobacter crescentus  
C;Species: Caulobacter crescentus  
C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C;Accession: C87552  
R;Nierman, W.C.; Feidiblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A;Title: Complete Genome Sequence of Caulobacter crescentus.  
A;Reference number: A87249; MUID:21173698; PMID:11259647  
A;Accession: C87552  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-433 <STO>  
A;Cross-references: GB:AE005673; NID:g13423989; PIDN:AAK24415.1; GSPDB:GN00148  
C;Genetics:  
A;Gene: CC2444  
C;Superfamily: Bacillus dihydroorotase; Bacillus dihydroorotase homology

Query Match 43.8%; Score 46; DB 2; Length 433;  
Best Local Similarity 40.0%; Pred. No. 22;  
Matches 8; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 DPAKGMSPPGFIVGEEGVLS 20  
DB 16 DPESGYDGGGVIVSEGVIT 35

RESULT 4  
S09804  
hypothetical protein UL41 - human cytomegalovirus (strain AD169)  
C;Species: human cytomegalovirus, human herpesvirus 5  
A;Note: host Homo sapiens (man)  
C;Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 28-Jul-2000  
C;Accession: S09804  
R;Chen, M.S.; Bankier, A.T.; Beck, S.; Bohni, R.; Brown, C.M.; Cerny, R.; Horsnell, T.;  
M.; Barrell, B.G.  
Curr. Top. Microbiol. Immunol. 154, 125-169, 1990  
A;Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus  
A;Reference number: S09749; MUID:90269039; PMID:2161319  
A;Accession: S09804  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-141 <CHE>  
A;Cross-references: EMBL:X17403; NID:g59591; PIDN:CAA35400.1; PID:e27240; PID:g1813967  
A;Note: this sequence was submitted to the EMBL Data Library, December 1989  
C;Superfamily: human cytomegalovirus hypothetical protein UL41

Query Match 42.9%; Score 45; DB 2; Length 141;  
Best Local Similarity 53.3%; Pred. No. 8.7;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 AKGMSPPGFIVGEEGV 17  
DB 76 AXDLPPFGYVRGG 90

RESULT 5  
H72501  
hypothetical protein APE1990 - Aeropyrum pernix (strain K1)  
C;Species: Aeropyrum pernix  
C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jun-2000  
C;Accession: H72501  
R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takaha  
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K  
DNA Res. 6, 83-101, 1999  
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr  
A;Reference number: A72450; MUID:99310339; PMID:10382966  
A;Accession: H72501  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-111 <KAW>  
A;Cross-references: DDBJ:AP000063; NID:g5105654; PIDN:BAAB1000.1; PID:g1044786; PID:g510  
A;Experimental source: strain K1  
C;Genetics:  
C;Superfamily: Aeropyrum pernix hypothetical protein APE1990

Query Match 41.9%; Score 44; DB 2; Length 111;  
Best Local Similarity 44.4%; Pred. No. 9.6;  
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 DPAKGMSPPGFIVGEEGV 18  
DB 85 EAARGVGAGFPVFDGV 102

RESULT 6  
A99238  
conserved hypothetical protein [imported] - Sulfolobus solfataricus  
C;Species: Sulfolobus solfataricus  
C;Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 15-Jun-2001  
C;Accession: A99238  
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-y  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.  
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001  
A;Description: Sulfolobus solfataricus complete genome.  
A;Reference number: A99139  
A;Accession: A99238  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-408 <KUR>  
A;Cross-references: GB:AB006641; NID:g13814052; PIDN:AAK41160.1; GSPDB:GN00155  
C;Genetics:  
A;Gene: SSO0872  
C;Superfamily: hypothetical protein AFI590

Query Match 41.9%; Score 44; DB 2; Length 408;  
Best Local Similarity 57.1%; Pred. No. 42;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 KGMSPPGFIVGEEGV 17  
DB 50 EGSTPPSGIVGEGK 63

RESULT 7  
H72502  
hypothetical protein APE1998 - Aeropyrum pernix (strain K1)  
C;Species: Aeropyrum pernix  
C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C;Accession: H72502  
R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takaha  
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K  
DNA Res. 6, 83-101, 1999  
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr  
A;Reference number: A72450; MUID:99310339; PMID:10382966  
A;Accession: H72502  
A;Status: preliminary  
A;Molecule type: DNA

A;Residues: 1-410 <KAW>  
 A;Cross-references: DDBJ:AF000063; NID:G5105654; PIDN:BA81008.1; PID:G5105696  
 A;Experimental source: strain K1  
 C;Genetics:  
 A;Gene: APE1998  
 C;Superfamily: hypothetical protein AF1590

Query Match 41.9%; Score 44; DB 2; Length 410;  
 Best Local Similarity 50.0%; Pred. No. 42;  
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 KGMSPGFIIVGREG 17  
 DB 47 QGSTPPSILVGERG 60

RESULT 8  
 F70628  
 probable pta protein - Mycobacterium tuberculosis (strain H37RV)  
 C;Species: Mycobacterium tuberculosis  
 C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
 C;Accession: F70628  
 R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A;Authors: Scars, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A;Reference number: A70500; MUID:98295997; PMID:9634230  
 A;Accession: F70628  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-690 <COL>  
 A;Cross-references: GB:Z84724; GB:AL123456; NID:G3261708; PIDN:CA806578.1; PID:G1817674  
 A;Experimental source: strain H37RV  
 C;Genetics:  
 A;Gene: pta  
 C;Superfamily: phosphate acetyltransferase pta

Query Match 41.9%; Score 44; DB 2; Length 690;  
 Best Local Similarity 40.0%; Pred. No. 76;  
 Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 DPAKGMSPGFIIVGREGVLS 20  
 DB 182 DALRFTPPSYVVPPELLS 201

RESULT 9  
 T02672  
 hypothetical protein R31449.3 - human (fragment)  
 C;Species: Homo sapiens (man)  
 C;Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 31-Dec-2000  
 C;Accession: T02672  
 R;Lamerdin, J.E.; McCready, P.M.; Skowronski, B.; Adamsen, A.W.; Burkhardt-Schultz, K.; C.J.; Dangnan, L.; Poundstone, P.; Christensen, M.; Georgescu, A.; Avila, J.; Liu, S.; A.P.; Quan, G.; Krommiller, B.; Arellano, A.; Montgomey, M.; Ow, D.; Nolan, M.  
 submitted to the EMBL Data Library, June 1998  
 A;Authors: Trong, S.; Kobayashi, A.; Olsen, A.S.; Carrano, A.V.  
 A;Description: Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a serine  
 A;Reference number: Z14696  
 A;Accession: T02672  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-813 <LAW>  
 A;Cross-references: EMBL:AC005175; NID:G3253116; PID:G3253120  
 C;Genetics:  
 A;Map position: 19  
 A;Intons: 93/2; 125/2; 174/1; 228/2; 266/3; 331/1; 372/1; 474/3; 637/1; 678/3  
 A;Note: R31449.3

Query Match 41.9%; Score 44; DB 2; Length 813;  
 Best Local Similarity 53.8%; Pred. No. 92;

Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 PAKGMSPGFIIVG 14  
 DB 675 PAAGIPPPGLLTG 687

RESULT 10  
 I55603  
 reduced hepatic glutathione transporter with canalicular features - rat  
 C;Species: Rattus norvegicus (Norway rat)  
 C;Date: 26-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 07-Feb-1997  
 C;Accession: I55603  
 R;Yi, J.R.; Lu, S.; Fernandez-Checa, J.; Kaplowitz, N.  
 J. Clin. Invest. 93, 1841-1845, 1994  
 A;Title: Expression cloning of a rat hepatic reduced glutathione transporter with canalicular features - rat  
 A;Reference number: I55603; MUID:94216535; PMID:8163683  
 A;Accession: I55603  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-835 <RES>  
 A;Cross-references: EMBL:U06845; NID:G501037; PID:G501038  
 C;Keywords: liver

Query Match 41.9%; Score 44; DB 2; Length 835;  
 Best Local Similarity 64.3%; Pred. No. 95;  
 Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 PAKGMSPGFIIVG 15  
 DB 727 PTKGALSPFIIVG 740

RESULT 11  
 S65775  
 homeotic protein Hox2b - maize  
 C;Species: Zea mays (maize)  
 C;Date: 14-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 11-Jan-2002  
 C;Accession: S65775; S72455  
 R;Klinge, B.; Ueberlacker, B.; Korfhage, C.; Werr, W.  
 Plant Mol. Biol. 30, 439-453, 1996  
 A;Title: ZmHox: a novel class of maize homeobox genes.  
 A;Reference number: S65774; MUID:96189260; PMID:8605297  
 A;Accession: S65775  
 A;Status: preliminary; nucleic acid sequence not shown  
 A;Molecule type: mRNA  
 A;Residues: 1-1539 <KL1>  
 A;Cross-references: EMBL:X89761

R;Werr, W.  
 submitted to the EMBL Data Library, July 1995  
 A;Reference number: S72455  
 A;Accession: S72455  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-379, 'S', 381-1539 <WER>  
 A;Cross-references: EMBL:X89761; NID:G1143706; PIDN:CAA61910.1; PID:G1143707  
 C;Genetics:  
 A;Gene: Hox2b  
 C;Superfamily: unassigned homeobox proteins; homeobox homology  
 C;Keywords: DNA binding; homeobox; nucleus; transcription regulation  
 F;802-858/Domain: homeobox homology <HOX1>  
 F;1040-1096/Domain: homeobox homology <HOX2>

Query Match 41.9%; Score 44; DB 2; Length 1539;  
 Best Local Similarity 53.8%; Pred. No. 1.9e+02;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 AKGMSPGFIIVG 15  
 DB 1217 ARGVGPFGYTFGE 1229

RESULT 12

S58884  
Ran-binding protein 2 - human  
N;Alternate names: giant nucleopore protein Nup358; nucleoporin Nup358; RanBP2 protein  
C;Species: Homo sapiens (man)  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
C;Accession: S58884; A57545  
R;Tokoyama, N.; Hayaishi, N.; Seki, T.; Pante, N.; Ohba, T.; Nishii, K.; Kuma, K.; Hayashihara, Y.  
Nature 376, 184-188, 1995.  
A;Title: A giant nucleopore protein that binds Ran/TC4.  
A;Reference number: S58884; MUID:95327194; PMID:7603572  
A;Accession: S58884  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 1-3224 <YOK>  
A;Cross-references: EMBL:D42063; NID:g924266; PIDN:BAA07662.1; PID:g1009337  
A;Experimental source: cell type B-lymphocyte  
R;Wu, J.; Matunis, M.J.; Kraemer, D.; Blobel, G.; Coutavas, E.  
J. Biol. Chem. 270, 14209-14213, 1995  
A;Title: Nup358, a cytoplasmically exposed nucleoporin with peptide repeats, Ran-GTP binding site, and a tetrapeptide repeat homologous to the Ran-binding protein.  
A;Reference number: A57545; MUID:95294031; PMID:7775481  
A;Accession: A57545  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-776,'R','R',778-783,'R',785-3224 <WUA>  
A;Cross-references: GB:L41840; NID:g857367; PIDN:AAC41758.1; PID:g857368  
A;Experimental source: cell line HeLa  
C;Genetics:  
A;Gene: GDB:RANBP2; NUP358  
A;Cross-references: GDB:4642756; OMIM:601181  
A;Map position: 2cen-2q13  
C;Function:  
C;Superfamily: nucleoporin Nup358; cyclophilin homology; tetrapeptide repeat homologous to Ran-binding protein  
C;Keywords: leucine zipper  
F;26-59/Domain: tetrapeptide repeat homology <TTL>  
F;60-93/Domain: tetrapeptide repeat homology <TT2>  
F;450-471/Domain: leucine zipper #status predicted <LEU>  
F;3063-3224/Domain: cyclophilin homology <CYP>

Query Match 41.9% Score 44; DB 1; Length 3224;  
Best Local Similarity 47.4%; Pred. No. 4.4e+02;  
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 2 PAKGSPPGFVGEGLVS 20  
| : | | | : | |  
Db 2505 PKAVSPKVFGESEVKS 2523

RESULT 13  
F81444  
probable integral membrane protein Cj0263 [imported] - Campylobacter jejuni (strain NCTC 8626)

C;Species: Campylobacter jejuni  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
C;Accession: F81444  
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Kesley, J.M.; Churcher, C.; Basham, D.; Chillingworth, C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrett, T.; et al.  
Nature 403, 665-668, 2000  
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyaluronate synthase and other virulence factors  
A;Reference number: A81250; MUID:20150912; PMID:10688204  
A;Accession: F81444  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-291 <PAR>  
A;Cross-references: GB:AL111168; NID:g967505; PIDN:CAB72731.1; PID:g596774  
A;Experimental source: serotype O2, strain NCTC 11168  
C;Genetics:  
A;Gene: CJ0263  
C;Superfamily: gufA protein

Query Match 41.0% Score 43; DB 2; Length 291;  
Best Local Similarity 53.3%; Pred. No. 41;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:31:13 ; Search time 0.767093 Seconds  
(without alignments)  
1357.597 Million cell updates/sec

Title: US-09-662-293-13

Perfect score: 105

Sequence: 1 DPXKGMSPPGFIVGEGVLS 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	45.7	440	1 T10B HUMAN	O14763 homo sapien
2	45	42.9	141	1 UL41_HCMVA	P16814 human cytom
3	45	42.9	887	1 ITH3_RAT	Q63416 rattus norv
4	44	41.9	232	1 YORL_TTV1	P19296 thermoprote
5	44	41.9	622	1 B1N4_MOUSE	Q920h1 mus muscucu
6	44	41.9	690	1 PTA_MYCTU	P96254 mycobacteri
7	44	41.9	3224	1 RB32_HUMAN	P49792 homo sapien
8	43	41.0	291	1 ZUPT_CAMEL	Q9pin2 campylobact
9	43	41.0	305	1 PPN1_ANASP	Q8yn19 anabaena sp
10	43	41.0	367	1 BET3_MESAU	O09029 mesocricetu
11	43	41.0	394	1 YE10_HAEIN	P44184 haemophilus
12	43	41.0	551	1 AAP2_NEUCR	O59942 neurospora
13	43	41.0	817	1 YG4A_YEAST	P46949 saccharomyc
14	42.5	40.5	646	1 INC9_CAEEL	P34542 caenorhabdi
15	42.5	40.5	3176	1 CA36_HUMAN	P12111 homo sapien
16	42	40.0	435	1 STCB_EMENI	Q13608 emericeila
17	42	40.0	677	1 SKD3_MOUSE	Q60649 mus muscucu
18	42	40.0	677	1 SKD3_RAT	Q9wrt2 rattus norv
19	42	40.0	770	1 LI00_ADE40	P11823 human adeno
20	42	40.0	886	1 ITH3_MESAU	P97280 mesocricetu
21	42	40.0	907	1 MOP_DESGI	Q46509 desulfocivib
22	42	40.0	1085	1 RBP2_BOVIN	P48820 bos taurus
23	41	39.0	234	1 ENGB_RALSO	Q8xv43 raietonia s
24	41	39.0	247	1 BRXB_CHICK	Q9ded6 gallus gall
25	41	39.0	263	1 CB22_ORISA	P12331 oryza sativ
26	41	39.0	266	1 CB21_ORISA	P12330 oryza sativ
27	41	39.0	369	1 LEU2_BUCUL	Q9agc6 buchnera ap
28	41	39.0	394	1 DXR_SYNY3	Q55663 synchocyst
29	41	39.0	442	1 LEU2_BUCUH	Q9evh7 buchnera ap
30	41	39.0	453	1 SR54_ECOLI	P07019 escherichia
31	41	39.0	707	1 SKD3_HUMAN	Q9A078 homo sapien
32	41	39.0	845	1 CLPC_CHLPN	Q928ae chlamydia p
33	41	39.0	854	1 CLPC_CHLTR	O84288 chlamydia t

#### ALIGNMENTS

##### RESULT 1

ID	T10B HUMAN	STANDARD;	PRT;	440 AA.
AC	O14763; O14720; O15508; O15517; O15531; Q9BVE0;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Tumor necrosis factor receptor superfamily member 10B precursor (Death receptor 5) (TNF-related apoptosis-inducing ligand receptor 2) (TRAIL receptor-2) (TRAIL-R2).			
GN	TNFRSF10B OR DR5 OR TRAILR2 OR TRICK2 OR KILLER OR ZTNFR9.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RP	(1)			
RP	SEQUENCE FROM N.A. (ISOFORM LONG), AND SEQUENCE OF N-TERMINUS.			
RC	TISSUE=Foreskin fibroblast;			
EX	MEDLINE=9745925; PubMed=9311998;			
RA	Walczak H., Degli-Esposti M.A., Johnson R.S., Smolak P.J., Waugh J.Y., Bolani N., Timour M.S., Gerhart M.J., Schooley K.A., Smith C.A., Goodwin R.G., Rauch C.T.;			
RA	"TRAIL-R2; a novel apoptosis-mediating receptor for TRAIL.;"			
RL	EMBO J. 16:5386-5397(1997).			
RP	(2)			
RP	SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT), ALTERNATIVE SPLICING, AND VARIANTS LEU-32 AND VAL-67.			
RP	MEDLINE=97431692; PubMed=9285725;			
RA	Screaton G.R., Mongkolsapaya J., Xu X.-N., Cowper A.E., McMichael A.J., Bell J.I.;			
RA	"TRICK2, a new alternatively spliced receptor that transduces the cytotoxic signal from TRAIL.;"			
RL	Curr. Biol. 7:693-696(1997).			
RP	(3)			
RP	SEQUENCE FROM N.A. (ISOFORM LONG), CHARACTERIZATION, AND VARIANTS LEU-32 AND VAL-67.			
RC	TISSUE=Liver, and Spleen;			
RA	MEDLINE=98039016; PubMed=9373179;			
RA	Schneider P., Bodmer J.-L., Thome M., Hofmann K., Holler N., Tschoopp J.;			
RA	"Characterization of two receptors for TRAIL.;"			
RL	FEBS Lett. 416:329-334(1997).			
RP	(4)			
RP	SEQUENCE FROM N.A. (ISOFORM SHORT).			
RC	TISSUE=Ovary;			
RA	MEDLINE=97467719; PubMed=9326928;			
RA	Wu G.S., Burns T.F., McDonald E.R. III, Jiang W., Meng R., Krantz I.D., Kao G., Gan D.D., Zhou J.Y., Muschel R., Hamilton S.R., Spinner N.B., Markowitz S., Wu G., el-Deiry W.S.;			
RA	"KILLER/DR5 is a DNA damage-inducible p53-regulated death receptor gene.;"			
RL	Nat. Genet. 17:141-143(1997).			
RP	(5)			
RP	SEQUENCE FROM N.A. (ISOFORM SHORT).			
RA	MEDLINE=97390508; PubMed=9242610;			
RA	Pan G., Ni J., Wei Y.-F., Yu G.-I., Gentz R., Dixit V.M.;			

O9pkas chlamydia m  
O9xik4 arabidopsis  
P29400 homo sapien  
O8r418 mus muscucu  
Q9up73 homo sapien  
P73069 synchocyst  
P51786 homo sapien  
O9kh57 mycobacteri  
P12942 mycobacteri  
P21160 mycobacteri  
P31952 mycobacteri  
Q06947 mycobacteri

34 41 39.0 870 1 CLPC\_CHLMU  
35 41 39.0 1058 1 U202\_ARATH  
36 41 39.0 1685 1 CAS4\_HUMAN  
37 41 39.0 1906 1 DICE\_MOUSE  
38 41 39.0 1912 1 DICE\_HUMAN  
39 40.5 38.6 342 1 Y48L\_SYNY3  
40 40.5 38.6 506 1 Z157\_HUMAN  
41 40 38.1 139 1 A85A\_MYCMB  
42 40 38.1 325 1 A85B\_MYCBO  
43 40 38.1 325 1 A85B\_MYCKA  
44 40 38.1 325 1 A85B\_MYCTU  
45 40 38.1 330 1 A85B\_MYCAV

RT "An antagonist decoy receptor and a death domain-containing receptor  
for TRAIL";  
RL Science 277:815-818(1997).  
RN [6]  
RP SEQUENCE FROM N.A. (ISOFORM SHORT), AND VARIANTS LEU-32 AND VAL-67.  
RX MEDLINE=97467318; PubMed=9325248;  
RA MacFarlane M., Ahmad M., Srinivasula S.M., Fernandes-Alnemri T.,  
Cohen G.M., Alnemri E.S.;  
RA "Identification and molecular cloning of two novel receptors for the  
cytotoxic ligand TRAIL";  
RL J. Biol. Chem. 272:25417-25420(1997).  
RN [7]  
RP SEQUENCE FROM N.A. (ISOFORM SHORT), AND VARIANTS LEU-32 AND VAL-67.  
RX MEDLINE=98090092; PubMed=9430227;  
RA Chaudhary P.M., Eby M., Jasmin A., Bookwalter A., Murray J., Hood L.;  
RA Chaudhary P.M., Eby M., Jasmin A., Bookwalter A., Murray J., Hood L.;  
RA FADD-dependent apoptosis and activate the NF-kappaB pathway";  
RT FADD-dependent apoptosis and activate the NF-kappaB pathway";  
RN Immunity 7:821-830(1997).  
RN [8]  
RP SEQUENCE FROM N.A. (ISOFORM SHORT), AND VARIANT LEU-32.  
RX MEDLINE=9730509; PubMed=9242611;  
RA Sheridan J.P., Marsters S.A., Pitti R.M., Gurney A., Skubatch M.,  
Baldwin D., Ramakrishnan L., Gray C.L., Baker K., Wood W.I.,  
Godard A.D., Godowski P., Ashkenazi A.;  
RA "Control of TRAIL-induced apoptosis by a family of signaling and decoy  
receptors";  
RL Science 277:818-821(1997).  
RN [9]  
RP SEQUENCE FROM N.A. (ISOFORM LONG), AND VARIANTS LEU-32 AND VAL-67.  
RX Arai T., Akiyama Y., Okabe S., Saito K., Iwai T., Yuasa Y.;  
RA "Genomic structure and mutation analyses of the DR5/TRAIL receptor 2  
gene in colorectal carcinoma";  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
RN [10]  
RP SEQUENCE FROM N.A. (ISOFORM SHORT).  
RX Cao X., Zhang W., Wan T.;  
RA Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
RN [11]  
RP SEQUENCE FROM N.A. (ISOFORM SHORT), AND VARIANTS LEU-32 AND VAL-67.  
RX Farrah T., Vu T., Gilbert T., Gross J., O'Hara P.;  
RA "Homo sapiens homolog of tumor necrosis factor receptor";  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
RN [12]  
RP SEQUENCE FROM N.A. (ISOFORM LONG), AND VARIANT LEU-32.  
RX TISSUE=Cervix;  
RL MEDLINE=2338257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,  
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
Bosak S.A., Morvan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
Villalon D.K., Muzny N.D., Sodergren E.J., Lu X., Gibbs R.A.,  
Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
Whiting M., Madan A., Young A.C., Shevchenko Y., Bonifard G.G.,  
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,  
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RA "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [13]  
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 54-183.  
RX MEDLINE=20017054; PubMed=10549288;  
RA Hymowitz S.G., Christinger H.W., Fuh G., Ultsch M., O'Connell M.,  
Kelley R.F., Ashkenazi A., de Vos A.M.;  
RA "Triggering cell death: the crystal structure of Apo2L/TRAIL in a  
complex with death receptor 5";

RL Mol. Cell 4:563-571(1999).  
RN [14]  
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 69-184.  
RX PubMed=10542098;  
RA Mongkolsapaya J., Grimes J.M., Chen N., Xu X.-N., Stuart D.I.,  
Jones E.Y., Screaton G.R.;  
RA "Structure of the TRAIL-DR5 complex reveals mechanisms conferring  
specificity in apoptotic initiation";  
RL Nat. Struct. Biol. 6:1048-1053(1999).  
CC -!- FUNCTION: Receptor for the cytotoxic ligand TNFSF10/TRAIL. The  
adaptor molecule FADD recruits caspase-8 to the activated  
receptor. The resulting death-inducing signaling complex (DISC)  
performs caspase-8 proteolytic activation which initiates the  
subsequent cascade of caspases (aspartate-specific cysteine  
proteases) mediating apoptosis. Promotes the activation of NF-  
kappaB.  
CC -!- SUBUNIT: Homotrimer. Can interact with TRADD and RIP.  
CC -!- SUBCELLULAR LOCATION: Type 1 membrane protein.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=Long; Synonyms=TRICK2B;  
CC IsoId=014763-1; Sequence=Displayed;  
CC Name=Short; Synonyms=TRICK2A;  
CC IsoId=014763-2; Sequence=VSP\_006490;  
CC -!- TISSUE SPECIFICITY: Widely expressed in adult and fetal tissues;  
very highly expressed in tumor cell lines such as HeLa S3, K562,  
HL-60, SW480, A549 and G361; highly expressed in heart, peripheral  
blood lymphocytes, liver, pancreas, spleen, thymus, prostate,  
ovary, uterus, placenta, testis, esophagus, stomach and throughout  
the intestinal tract; not detectable in brain.  
CC -!- INDUCTION: TNFRSF10B is regulated by the tumor suppressor p53.  
CC -!- DISEASE: Defects in TNFRSF10B may be a cause of squamous cell  
carcinoma of the head and neck (HNSCC) [MIM:601400].  
CC -!- SIMILARITY: Contains 3 TNFR-Cys repeats.  
CC -!- SIMILARITY: Contains 1 death domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----  
DR EMBL; AF016849; AAC51778.1; -  
DR EMBL; AF018657; AAB70577.1; -  
DR EMBL; AF018658; AAB70578.1; -  
DR EMBL; AF016266; AAB81180.1; -  
DR EMBL; AF022386; AAB71949.1; -  
DR EMBL; AF016268; AAB67103.1; -  
DR EMBL; AF020501; AAB71412.1; -  
DR EMBL; AF016268; AAC01565.1; -  
DR EMBL; AF012535; AAB67103.1; -  
DR EMBL; AB014718; BAA33723.1; -  
DR EMBL; AB014710; BAA33723.1; JOINED.  
DR EMBL; AB014711; BAA33723.1; JOINED.  
DR EMBL; AB014712; BAA33723.1; JOINED.  
DR EMBL; AB014713; BAA33723.1; JOINED.  
DR EMBL; AB014714; BAA33723.1; JOINED.  
DR EMBL; AB014715; BAA33723.1; JOINED.  
DR EMBL; AB014716; BAA33723.1; JOINED.  
DR EMBL; AB014717; BAA33723.1; JOINED.  
DR EMBL; AF153687; AAF75587.1; -  
DR EMBL; AF192548; AAF07175.1; -  
DR EMBL; BC001281; AAH01281.1; -  
DR FDB; 1D0G; 22-OCT-99.  
DR FDB; 1D4V; 01-NOV-99.  
DR Genew; HGNC:11905; TNFRSF10B.  
DR MIM; 603612; -  
DR MIM; 601400; -  
DR GO; GO:0016021; C:integral to membrane; IC.  
DR GO; GO:0016506; F:apoptosis activator activity; NAS.  
DR GO; GO:0008656; F:caspase activator activity; NAS.

```

Query Match      45.7%; Score 48; DB 1; Length 440;
Best Local Similarity 43.8%; Pred. No. 7.8;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY      2 PAKGMSPPGFIIVGEEG 17
Db      76 PSEGLCPGHHSIEDG 91

RESULT 2
UL41_HCMVA
ID      UL41_HCMVA      STANDARD;      PRT;      141 AA.
AC      P168T4;
DT      01-AUG-1990 (Rel. 15, Created)
DT      01-AUG-1990 (Rel. 15, Last sequence update)
DT      01-FEB-1991 (Rel. 17, Last annotation update)
DE      Hypothetical protein UL41.
GN      UL41.
OS      Human cytomegalovirus (strain AD169).
OC      Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC      Betaherpesvirinae; Cytomegalovirus.
ON      NCBI_TaxID=10360;
RX      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90269039; PubMed=2161319;
RA      Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,
RA      Horsnell T., Hutchison C.A. III, Kuzarides T., Martignetti J.A.,
RA      Freddie E., Satchell S.C., Tomlinson P., Weston K.M., Barrell B.G.;
RT      "Analysis of the protein-coding content of the sequence of human
RT      cytomegalovirus strain AD169.";
RL      Curr. Top. Microbiol. Immunol. 154:125-169(1990).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; X83231; CAA58233.1; -
CC      InterPro; IPR006587; VIT.
CC      InterPro; IPR002035; VWF_A.
CC      Pfam; PF000092; vwa; 1.
CC      SMART; SM00609; VIT; 1.
CC      SMART; SM00327; VWA; 1.
CC      PROSITE; PS0234; VWF_A; 1.
CC      KW      Serine protease inhibitor; Repeat; Signal; Multigene family;
CC      Glycoprotein.
CC      FT      SIGNAL      1      21      POTENTIAL.
CC      FT      PROPEP      22      33      BY SIMILARITY.
CC      FT      CHAIN      34      647      INTER-ALPHA-TRYPsin INHIBITOR HEAVY CHAIN
CC      FT      PROPEP      648      887      H3.
CC      FT      DOMAIN      282      442      VMPA.
CC      FT      CARBOHYD      91      91      N-LINKED (GLCNAC. .) (POTENTIAL).
CC      FT      CARBOHYD      580      580      N-LINKED (GLCNAC. .) (POTENTIAL).
CC      FT      BINDING      647      647      CHONDROITIN 4-SULFATE, CROSS-LINK SITE
CC      FT      BINDING      647      647      (BY SIMILARITY).
CC      SQ      SEQUENCE      887 AA; 99097 MW; 3B9F0FF96D514096 CRC64;

Query Match      42.9%; Score 45; DB 1; Length 887;
Best Local Similarity 47.1%; Pred. No. 48;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY      1 DPAKMSPPGFIIVGEEG 17
Db      678 DPVTGIATVGTGIIIEGK 694

RESULT 4
YORL_TTVI
ID      YORL_TTVI      STANDARD;      PRT;      232 AA.
AC      P19296;
DT      01-NOV-1990 (Rel. 16, Created)
DT      01-NOV-1990 (Rel. 16, Last sequence update)
DT      01-NOV-1990 (Rel. 16, Last annotation update)
DE      Hypothetical 26.9 kDa protein.
OS      Thermoproteus tenax virus 1 (strain KRA1) (TTV1).
OC      Viruses; dsDNA viruses, no RNA stage; Lipothrixviridae;
OC      Lipothrixvirus.
OX      NCBI_TaxID=10480;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Neumann H.;
RL      Submitted (MAR-1989) to the EMBL/GenBank/DBJ databases.
CC      -----

```



CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; X14855; CAA32992.1; -;  
 CC SEQUENCE 232 AA; 26862 MW; 0ABC274AF657CA9D CRC64;  
 CC

Query Match 41.9%; Score 44; DB 1; Length 232;  
 Best Local Similarity 54.5%; Pred. No. 16;  
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 8 PGFIVGEGV 18  
 |||:|:|:  
 Db 176 PGQLGDEGI 186

RESULT 5  
 BIN4\_MOUSE STANDARD; PRT; 622 AA.  
 AC Q920H1;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE WD-repeat protein BIN4.  
 GN BING4.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CX NCBI\_TaxID=10090;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=129/SvJ;  
 RA Bowen L., Qin S., Madan A., Loretz C., Hall J., James R., Dors M.,  
 RA Shaffer T., Abbasi N., Ratcliffe A., Dickhoff R., Lasky S., Hood L.,  
 RA "Sequence of the mouse major histocompatibility complex class II  
 RT region."  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Olfactory epithelium;  
 RX MEDLINE=23388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RL human and mouse cDNA sequences."  
 CC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 CC -1- SIMILARITY: Contains 5 WD repeats.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; AF110520; AAC97976.1; -;  
 CC EMBL; AF100956; AAC69896.1; -;  
 CC EMBL; BC046977; AAH46977.1; -;  
 CC MGB; MGI:1931871; BING4.  
 CC InterPro; IPR001680; WD40.  
 CC Pfam; PF00400; WD40; 1.  
 CC SMART; SM00320; WD40; 3.  
 CC PROSITE; PS00678; WD\_REPEATS\_1; FALSE\_NEG.  
 CC PROSITE; PS0082; WD\_REPEATS\_2; 1.  
 CC PROSITE; PS0294; WD\_REPEATS\_REGION; 1.

KW Repeat; WD repeat. 233 WD 1.  
 FT REPEAT 192 233 WD 2.  
 FT REPEAT 234 271 WD 3.  
 FT REPEAT 314 353 WD 4.  
 FT REPEAT 356 395 WD 5.  
 FT REPEAT 398 435  
 SQ SEQUENCE 622 AA; 69048 MW; B1BF9A36AC612612 CRC64;

Query Match 41.9%; Score 44; DB 1; Length 622;  
 Best Local Similarity 77.8%; Pred. No. 46;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 9 PGFIVGEGV 17  
 |||:|:|:  
 Db 156 PGFLVGEDG 164

RESULT 6  
 PTA\_MYCTU STANDARD; PRT; 690 AA.  
 AC P96254;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Phosphate acetyltransferase (EC 2.3.1.8) (Phosphotransacetylase).  
 GN PTA OR RV0408 OR MT0421 OR MTCY22G10.04.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 CX NCBI\_TaxID=1773;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544 (1998).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RX MEDLINE=22206494; PubMed=12218036;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;  
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains."  
 RL J. Bacteriol. 184:5479-5490 (2002).  
 CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + phosphate = CoA + acetyl  
 CC phosphate  
 CC -1- PATHWAY: Conversion of acetate to acetyl-CoA; second step.

```

CC CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; CONTAINS A DOMAIN WITH SOME
CC CC SIMILARITIES WITH COBQ/BIOD.
CC CC -!- SIMILARITY: In the C-terminal section; belongs to the phosphate
CC CC acetyltransferase and butyryltransferase family.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; Z84724; CAB06578.1; -.
CC CC EMBL; AE006946; AAK44645.1; -.
CC CC PIR; F70628; F70628.
CC CC TIGR; MT0421; -.
CC CC Tuberculist; RV0408; -.
CC CC InterPro; IPR004614; Pta.
CC CC InterPro; IPR002505; PTA_PTB.
CC CC Pfam; PF01515; PTA_PTB; 1.
CC CC TIGRFAMs; TIGR00651; pta; 1.
CC CC Transferase; Acyltransferase; Complete proteome.
CC CC FT DOMAIN 365 690 PHOSPHATE ACETYLTRANSFERASE.
CC CC SQ SEQUENCE 690 AA; 72948 MW; C01C412AF2810CCE CRC64;
CC CC -----
CC CC Query Match 41.9%; Score 44; DB 1; Length 690;
CC CC Best Local Similarity 40.0%; Pred. No. 52;
CC CC Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;
CC CC -----
CC CC QY 1 DPAKMSPPGFIYVGEGLS 20
CC CC Db 182 DALRETPPSYVPEPLLS 201
CC CC -----
CC CC RESULT 7
CC CC ID RBP2_HUMAN STANDARD; PRT; 3224 AA.
CC CC AC P49792; Q15280;
CC CC DT 01-OCT-1996 (Rel. 34, Created)
CC CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
CC CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC CC DE Ran-binding protein 2 (RanBP2) (Nuclear pore complex protein Nup358)
CC CC DE (Nucleoporin Nup358) (358 kDa nucleoporin) (P270).
CC CC GN RANBP2 OR NUP358.
CC CC OS Homo sapiens (Human).
CC CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC CC OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CC CC OX NCBI_TaxID=9606;
CC CC RN [1]
CC CC RP SEQUENCE FROM N.A.
CC CC RX MEDLINE=95294031; PubMed=7775481;
CC CC RA Wu J., Matunis M.J., Kraemer D., Blobel G., Coutavas E.;
CC CC RT "Nup358, a cytoplasmically exposed nucleoporin with peptide repeats,
CC CC RT Ran-GTP binding sites, zinc fingers, a cyclophilin A homologous
CC CC RT domain, and a leucine-rich region.";
CC CC RL J. Biol. Chem. 270:14209-14213(1995).
CC CC RN [2]
CC CC RP SEQUENCE FROM N.A.
CC CC RC TISSUE=Blood;
CC CC RX MEDLINE=95327194; PubMed=7603572;
CC CC RA Yokoyama N., Hayashi N., Seki T., Nishii K., Hayashida T.,
CC CC RA Kuma K.I., Miyata T., Fukui M., Nishimoto T., Pante N., Aebi U.;
CC CC RT "A giant nucleopore protein that binds Ran/TC4.";
CC CC RL Nature 376:184-188(1995).
CC CC RN [3]
CC CC RP X-RAY CRYSTALLOGRAPHY (2.96 ANGSTROMS) OF 1171-1304.
CC CC RX MEDLINE=99176415; PubMed=10078529;
CC CC RA Vetter I.R., Nowak C., Nishimoto T., Kuhlmann J., Wittinghofer A.;
CC CC RT "Structure of a Ran-binding domain complexed with Ran bound to a GTP
CC CC RT analogue: implications for nuclear transport.";
CC CC RL Nature 398:39-46(1999).

```

```

CC CC -!- FUNCTION: Involved in transport factor (Ran-GTP, karyopherin)-
CC CC mediated protein import via the F-G repeat-containing domain which
CC CC acts as a docking site for substrates. Could also have isomerase
CC CC or chaperone activity and may bind RNA or DNA. Component of the
CC CC nuclear export pathway. Specific docking site for the nuclear
CC CC export factor exportin-1.
CC CC -!- SUBUNIT: Forms a tight complex in association with RANBP1 and the
CC CC ubiquitin-conjugating enzyme E2 (UBC9) (By similarity).
CC CC -!- SUBCELLULAR LOCATION: Nuclear pore complex. Cytoplasmic filaments.
CC CC -!- DOMAIN: Contains P-X-P-G repeats.
CC CC -!- SIMILARITY: Contains 4 RANBD1 domains.
CC CC -!- SIMILARITY: Contains 8 RANBP2-type zinc fingers.
CC CC -!- SIMILARITY: Contains 1 cyclophilin-like PPIase domain.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; L41840; AAC41758.1; -.
CC CC EMBL; D42063; BAA07662.1; -.
CC CC PIR; S58884; S58884.
CC CC PDB; 1RRP; 18-MAY-99.
CC CC Genew; HGNC:9848; RANBP2.
CC CC MIM; 601181; -.
CC CC GO; GO:0005643; C:nuclear pore; TAS.
CC CC GO; GO:0008336; F:RAN protein binding; TAS.
CC CC GO; GO:0006606; P:protein-nucleus import; TAS.
CC CC InterPro; IPR002130; CSA_PPIase.
CC CC InterPro; IPR000697; EVH1.
CC CC InterPro; IPR000156; Ran_BP1.
CC CC InterPro; IPR008941; TPR-like.
CC CC InterPro; IPR001440; TPR.
CC CC InterPro; IPR001876; Znf_RanGDP.
CC CC Pfam; PF00160; pro_isomerase; 1.
CC CC Pfam; PF00638; Ran_BP1; 4.
CC CC Pfam; PF00515; TPR; 1.
CC CC Pfam; PF00641; zf-RanBP; 8.
CC CC PRINTS; PR00153; CSA_PPIASMEASE.
CC CC SMART; SM00360; RANBD; 4.
CC CC SMART; SM00547; Znf_RBZ; 8.
CC CC PROSITE; PS00170; CSA_PPIASE_1; 1.
CC CC PROSITE; PS00072; CSA_PPIASE_2; 1.
CC CC PROSITE; PS00196; RANBD1; 4.
CC CC PROSITE; PS01358; ZF_RANBP2_1; 8.
CC CC PROSITE; PS0199; ZF_RANBP2_2; 8.
CC CC Nuclear protein; Transport; Repeat; Zinc-finger; Isomerase; Rotamase;
CC CC 3D-structure; Polymorphism.
CC CC DOMAIN 1171 1307 RANBD1 1.
CC CC ZN_FING 1351 1381 RANBP2-TYPE 1.
CC CC ZN_FING 1415 1444 RANBP2-TYPE 2.
CC CC ZN_FING 1479 1508 RANBP2-TYPE 3.
CC CC ZN_FING 1543 1572 RANBP2-TYPE 4.
CC CC ZN_FING 1606 1635 RANBP2-TYPE 5.
CC CC ZN_FING 1665 1694 RANBP2-TYPE 6.
CC CC ZN_FING 1724 1753 RANBP2-TYPE 7.
CC CC ZN_FING 1781 1810 RANBP2-TYPE 8.
CC CC DOMAIN 2012 2148 RANBD1 2.
CC CC DOMAIN 2309 2445 RANBD1 3.
CC CC DOMAIN 2911 3046 RANBD1 4.
CC CC DOMAIN 3067 3223 PPIASE, CYCLOPHILIN-TYPE.
CC CC VARIANT 1892 1892 P -> A (in dbSNP:12770).
CC CC CONFLICT 777 777 R -> H (IN REF. 2).
CC CC CONFLICT 784 784 R -> K (IN REF. 2).
CC CC STRAND 1191 1204
CC CC STRAND 1211 1224
CC CC STRAND 1231 1235
CC CC TURN 1237 1239
CC CC STRAND 1242 1244

```

```
FT STRAND 1247 1247
FT STRAND 1255 1255
FT TURN 1258 1259
FT TURN 1261 1262
FT STRAND 1263 1270
FT TURN 1272 1273
FT STRAND 1277 1284
FT TURN 1288 1300
FT TURN 1301 1302
SQ SEQUENCE 3224 AA; 548214 MW; 54E78412C96A3C63 CRC64;

Query Match 41.9%; Score 44; DB 1; Length 3224;
Best Local Similarity 47.4%; Pred. No. 2.7e+02;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 2 PAKGMSPPGFVGBEGLVLS 20
DB 2505 PKAVVSPPKFVFGSESVKS 2523

RESULT 8
ZUPT CAMJE STANDARD; PRT; 291 AA.
ID ZUPT CAMJE STANDARD; PRT; 305 AA.
AC Q9PIN2;
DT 28-FEB-2003 (Rel. 41, Created)
DE 10-OCT-2003 (Rel. 42, Last sequence update)
DE Zinc transporter zupT.
GN ZUPT OR CU0263.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
CC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kelsey J.M., Churcher C.,
RA Bagham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagers K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrall B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RL Nature 403:665-668(2000).
CC
CC -!- FUNCTION: Mediates zinc uptake. May also transport other divalent
CC cations (by similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: Belongs to the ZIP transporter (TC 2.A.5) family. ZupT
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL139074; CAB72731.1; -.
CC PIR; F81444; F81444.
CC HAMAP; MF_00548; -.
CC InterPro; IPR003689; Zn_transp_Zip.
CC Pfam; PF02535; Zip; 1.
CC Transmem 8
CC Transmem 28
CC Transmem 39
CC Transmem 74
CC Transmem 94
CC Transmem 147
CC Transmem 167
CC Transmem 174
CC Transmem 194
CC Transmem 209
CC Transmem 229
CC Transmem 233
CC Transmem 253
CC Transmem 271
CC Transmem 291

Query Match 41.0%; Score 43; DB 1; Length 305;
Best Local Similarity 30.8%; Pred. No. 31;
Matches 8; Conservative 6; Mismatches 4; Indels 8; Gaps 1;

QY 2 PAKGMSPPGFVGBEGLVLS 19
DB 56 PIDGLTPPGFSEMKFAIVLGDDGTV 81

RESULT 10
BET3_MESAU STANDARD; PRT; 367 AA.
ID BET3_MESAU
AC O09029;
DT 30-MAY-2000 (Rel. 39, Created)
```

30-MAY-2000 (Rel. 39, Last sequence update)  
15-MAR-2004 (Rel. 43, Last annotation update)  
BETA3 protein.  
Mesocricetus auratus (Golden hamster).  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
Mesocricetus.  
NCBI\_TaxID=10036;  
[1]  
SEQUENCE FROM N.A.  
MEDLINE=96140430; PubMed=8552091;  
Peyton M., Stelrecht C.M.M., Naya F.J., Huang H.-P., Samora P.J.,  
Tsai M.-J.;  
"BETA3, a novel helix-loop-helix protein, can act as a negative  
regulator of BETA2 and MyoD-responsive genes.";  
Mol. Cell. Biol. 16:626-633(1996).  
-!- FUNCTION: Inhibits DNA binding of TCF3 (E47) homodimers and TCF3  
NeuroD and MyoD-responsive genes, probably by heterodimerization  
with class A basic helix-loop-helix factors. Despite the presence  
of an intact basic domain, does not bind to DNA.  
-!- SUBUNIT: HETERO-DIMER WITH OTHER BHLH PROTEINS, LIKE TCF3 (E47).  
-!- SUBCELLULAR LOCATION: Nuclear (Potential).  
-!- TISSUE SPECIFICITY: Kidney, lung, brain and pancreas (insulinoma).  
-!- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; S80870; AAB50691.1; -  
TRANSFAC; T01674; -  
InterPro; IPR001092; HLH\_basic.  
Pfam; PF00010; HLH; 1.  
SMART; SM00353; HLH; 1.  
PROSITE; PS50888; HLH; 1.  
KW Nuclear protein; Transcription regulation; Repressor.  
FT DOMAIN 11 14 POLY-ALA.  
FT DOMAIN 58 62 POLY-SER.  
FT DOMAIN 83 99 POLY-GLY.  
FT DOMAIN 174 179 POLY-GLY.  
FT DOMAIN 204 217 POLY-GLY.  
FT DNA BIND 229 240 BASIC DOMAIN.  
FT DOMAIN 241 282 HELIX-LOOP-HELIX MOTIF.  
FT DOMAIN 311 319 POLY-ALA.  
SQ SEQUENCE 367 AA; 35905 MW; 6CAB9AFF96E85F77 CRC64;  
Query Match 41.0%; Score 43; DB 1; Length 367;  
Best Local Similarity 57.1%; Pred. No. 38;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
QY 5 GNSPPGFTVGGGV 18  
|:|:|:|:|:|:|  
Db 99 GVSVPGLLVGSAGV 112  
RESULT 11  
YE10 HAEIN  
ID YE10 HAEIN STANDARD; PRT; 394 AA.  
AC P44184;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Protein H11410.  
GN H11410.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OC NCBI\_TaxID=727;  
-----  
[1]  
SEQUENCE FROM N.A.  
STRAIN=RD / KW20 / ATCC 51907;  
MEDLINE=95350630; PubMed=7542800;  
Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
McKenny K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
Pine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
Venter J.C.;  
"Whole-genome random sequencing and assembly of Haemophilus influenzae  
Rd.";  
Science 269:496-512(1995).  
[2]  
IDENTIFICATION BY MASS SPECTROMETRY.  
RX MEDLINE=20137488; PubMed=10675023;  
Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,  
Gray C., Fountoulakis M.;  
"Two-dimensional map of the proteome of Haemophilus influenzae.";  
Electrophoresis 21:411-429(2000).  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; U32820; AAC23058.1; -  
PIR; E64028; E64028.  
TIGR; H11410; -  
InterPro; IPR006701; G2P.  
InterPro; IPR006437; Phage\_term\_2.  
Pfam; PF04466; G2P; 1.  
TIGRFAMS; TIGR01547; phage\_term\_2; 1.  
KW Complete proteome.  
SQ SEQUENCE 394 AA; 44782 MW; D4217986B6C772 CRC64;  
Query Match 41.0%; Score 43; DB 1; Length 394;  
Best Local Similarity 60.0%; Pred. No. 41;  
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
QY 3 AKGWSPPGFTVGGGV 17  
|:|:|:|:|:|:|  
Db 186 AKGKVKVGFVADVG 200  
RESULT 12  
AAP2 NEUCR  
ID AAP2 NEUCR STANDARD; PRT; 551 AA.  
AC O59942;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Amino acid permease 2.  
GN AAP-2 OR AAP2.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Margolles-Clark E., Bowman B.J.;  
Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
RL -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
CC -!- SIMILARITY: Belongs to the amino acid permease family.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL; AF053231; AAC08355.1; --  
CC InterPro; IPR002293; AA/permease1.  
CC InterPro; IPR004756; AA/permease.  
CC InterPro; IPR004840; AAC/permease.  
CC InterPro; IPR004841; Permease\_region.  
CC Pfam; PF00324; aa\_permeases; 1.  
CC TIGRFAMs; TIGR00907; 2A0304; 1.  
CC PROSITE; PS00218; AMINO-ACID PERMEASE 1; 1.  
CC Transport; Amino-acid transport; Transmembrane.  
KW TRANSMEM 66 86 POTENTIAL.  
FT TRANSMEM 90 110 POTENTIAL.  
FT TRANSMEM 139 159 POTENTIAL.  
FT TRANSMEM 188 208 POTENTIAL.  
FT TRANSMEM 214 234 POTENTIAL.  
FT TRANSMEM 255 275 POTENTIAL.  
FT TRANSMEM 301 321 POTENTIAL.  
FT TRANSMEM 347 367 POTENTIAL.  
FT TRANSMEM 399 419 POTENTIAL.  
FT TRANSMEM 424 444 POTENTIAL.  
FT TRANSMEM 464 484 POTENTIAL.  
FT TRANSMEM 496 516 POTENTIAL.  
SQ SEQUENCE 551 AA; 59869 MW; C812C646B82F9ADF CRC64;

Query Match 41.0%; Score 43; DB 1; Length 551;  
Best Local Similarity 70.0%; Pred. No. 58;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 PAKGMSPPPGF 11  
Db 252 PSSGWSPPGPF 261

## RESULT 13

YGA4 YEAST  
ID YG4A YEAST STANDARD; PRT; 817 AA.  
AC P46949;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical 90.8 kDa protein in RRP41-SNG1 intergenic region.  
GN YGR196C OR G7589.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C;  
RX MEDLINE=96076633; PubMed=7502584;  
RA Guerreiro P., Maia e Silva A., Barreiros T., Arroyo J.,  
RA Garcia-Gonzalez M., Garcia-Saez M.I., Rodrigues-Pousada C.,  
RA Nombela C.  
RT "The complete sequence of a 9000 bp fragment of the right arm of  
RT Saccharomyces cerevisiae chromosome VII contains four previously  
RT unknown open reading frames";  
RL Yeast 11:1087-1091(1995).  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL; X82775; CAA58019.1; --  
CC EMBL; Z72981; CAA97222.1; --

DR PIR; S53919; S53919.  
DR GerMOnline; 141508; --  
DR SGD; S0003428; FYV8.  
DR GO; GO:0005737; Cytoplasm; IDA.  
KW Hypothetical protein.  
SQ SEQUENCE 817 AA; 90797 MW; B52C5D659D63BEBB CRC64;

Query Match 41.0%; Score 43; DB 1; Length 817;  
Best Local Similarity 36.8%; Pred. No. 89;  
Matches 7; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 DPAKGMSPPGFIVGEGVL 19  
Db 547 DTANKAPGVVIDSNGKL 565

## RESULT 14

YNC9 CAEEL  
ID YNC9 CAEEL STANDARD; PRT; 646 AA.  
AC P34542; P34543;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical protein R05D3.9 in chromosome III precursor.  
GN R05D3.9/R05D3.10.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,  
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,  
RA Latreille P., Lightning J., Lloyd C., Morimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,  
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,  
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
RA Woldman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans";  
RL Nature 368:32-38(1994).  
RN [2]  
RP REVISIONS.

RA Waterston R.;  
RA Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL; L07144; AA21444.2; --  
CC PIR; S44863; S44863.  
CC PIR; S44865; S44865.  
CC WormPep; R05D3.9; CE31049.  
KW Hypothetical protein; signal.  
FT SIGNAL 1 29 POTENTIAL.  
FT CHAIN 30 646 HYPOTHETICAL PROTEIN R05D3.9.  
SQ SEQUENCE 646 AA; 72631 MW; 887932547DB5B141 CRC64;

Query Match 40.5%; Score 42.5; DB 1; Length 646;  
Best Local Similarity 58.8%; Pred. No. 82;  
Matches 10; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

QY 2 PAKGSPGPG-FIVGEG 17  
DB 42 PPSGSPGPGKFPDGE 58

RESULT 15  
CA36\_HUMAN STANDARD; PRT; 3176 AA.  
ID CA36\_HUMAN STANDARD; PRT; 3176 AA.  
AC P12111; Q16501;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Collagen alpha 3(VI) chain precursor.  
GN COL6A3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Fibroblast;  
RX MEDLINE=90151612; PubMed=1699238;  
RA Chu M.-L., Zhang R.-Z., Pan T.-C., Stokes D., Conway D., Kuo H.-J.,  
RA Glanville R., Mayer U., Mann K., Deutzmann R., Timpl R.;  
RA "Mosaic structure of globular domains in the human type VI collagen  
alpha 3 chain: similarity to von Willebrand factor, fibronectin,  
RT actin, salivary proteins and aprotinin type protease inhibitors.";  
RL ENBO J. 9:385-393(1990).  
RN [2]  
RP REVISIONS.  
RX Chu M.-L.;  
RA Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 2038-2373 FROM N.A.  
RX MEDLINE=89066644; PubMed=3198591;  
RA Chu M.-L., Conway D., Pan T.-C., Baldwin C., Mann K., Deutzmann R.,  
RA Timpl R.;  
RA "Amino acid sequence of the triple-helical domain of human collagen  
type VI";  
RT J. Biol. Chem. 263:18601-18606(1988).  
RN [4]  
RP SEQUENCE OF 2092-2157 FROM N.A.  
RX MEDLINE=88029444; PubMed=3665927;  
RA Chu M.-L., Mann K., Deutzmann R., Pribula-Conway D.,  
RA Hsu-Chen C.-C., Bernard M.P., Timpl R.;  
RA "Characterization of three constituent chains of collagen type VI by  
RT peptide sequences and cDNA clones";  
RL Eur. J. Biochem. 168:309-317(1987).  
RN [5]  
RP SEQUENCE OF 2092-2151 FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=88161046; PubMed=3348212;  
RA Weil D., Mattei M.-G., Passage E., van Cong N., Pribula-Conway D.,  
RA Mann K., Deutzmann R., Timpl R., Chu M.-L.;  
RA "Cloning and chromosomal localization of human genes encoding the  
RT three chains of type VI collagen";  
RL Am. J. Hum. Genet. 42:435-445(1988).  
RN [6]  
RP SEQUENCE OF 32-236 FROM N.A., AND ALTERNATIVE SPLICING.  
RX MEDLINE=93054780; PubMed=1339440;  
RA Zanussi S., Doliana R., Segat D., Bonaldo P., Colombatti A.;  
RA "The human type VI collagen gene. mRNA and protein variants of the  
RT alpha 3 chain generated by alternative splicing of an additional 5-end  
RT exon";  
RL J. Biol. Chem. 267:24082-24089(1992).  
RN [7]  
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 3107-3164.  
RX MEDLINE=95182468; PubMed=7533217;  
RA Arnoux B., Merigault K., Saludjian P., Norris K., Bjoern S.,  
RA Olsen O., Petersen L., Ducruix A.;  
RA "The 1.6 A structure of Kunitz-type domain from the alpha 3 chain of  
RT human type VI collagen";  
RL J. Mol. Biol. 246:609-617(1995).  
RN [8]  
RP STRUCTURE BY NMR OF 3102-3164.  
RX MEDLINE=96398604; PubMed=8805527;  
RA Zwickstetter M., Czisch M., Mayer U., Chu M.-L., Zinth W., Timpl R.,  
RA Holak T.A.;  
RA "Structure and multiple conformations of the Kunitz-type domain from  
RT human type VI collagen alpha3(VI) chain in solution.";  
RL Structure 4:195-209(1996).  
RN [9]  
RP STRUCTURE BY NMR OF 3107-3164.  
RX MEDLINE=97410331; PubMed=9265624;  
RA Soerensen M.D., Bjoern S., Norris K., Olsen O., Petersen L.,  
RA James T.L., Led J.J.;  
RA "Solution structure and backbone dynamics of the human alpha3-chain  
RT type VI collagen C-terminal Kunitz domain.";  
RL Biochemistry 36:10439-10450(1997).  
RN [10]  
RP DISEASE.  
RX MEDLINE=21987636; PubMed=11992252;  
RA Demir E., Sabatelli P., Allamand V., Ferreira A., Moghadaszadeh B.,  
RA Makrelouf M., Topaloglu H., Schenke B., Marlin L., Guicheney P.;  
RA "Mutations in COL6A3 cause severe and mild phenotypes of Ullrich  
RT congenital muscular dystrophy.";  
RL Am. J. Hum. Genet. 70:1446-1458(2002).  
RN [11]  
RP VARIANT BM GLU-1679, AND VARIANT HIS-2831.  
RX MEDLINE=98204804; PubMed=9536084;  
RA Pan T.-C., Zhang R.-Z., Pericak-Vance M.A., Tandan R., Fries T.,  
RA Stajich J.M., Viles K., Vance J.M., Chu M.-L., Speer M.C.;  
RA "Missense mutation in a von Willebrand factor type A domain of the  
RT alpha 3(VI) collagen gene (COL6A3) in a family with Bethlem  
RT myopathy";  
RL Hum. Mol. Genet. 7:807-812(1998).  
RN [12]  
RP FUNCTION: Collagen VI acts as a cell-binding protein.  
CC - SUBUNIT: Trimers composed of three different chains: alpha 1(VI),  
CC alpha 2(VI), and alpha 3(VI).  
CC - ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1;  
CC IsoId=PI2111-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=PI2111-2; Sequence=VSP 001172;  
CC - PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC - DISEASE: Defects in COL6A3 are a cause of Bethlem myopathy (BM)  
CC [MIM:158810]. BM is a rare autosomal dominant proximal myopathy  
CC characterized by early childhood onset (complete penetrance by the  
CC age of 5) and joint contractures most frequently affecting the  
CC elbows and ankles.  
CC - DISEASE: Defects in COL6A3 are a cause of Ullrich congenital  
CC muscular dystrophy (UCMD) [MIM:254090]; also known as Ullrich  
CC scleroatonic muscular dystrophy. UCMD is an autosomal recessive  
CC congenital myopathy characterized by muscle weakness and multiple  
CC joint contractures, generally noted at birth or early infancy. The  
CC clinical course is more severe than in Bethlem myopathy.  
CC - SIMILARITY: Contains 1 BPTI/Kunitz inhibitor domain.  
CC - SIMILARITY: Contains 1 fibronectin type III domain.  
CC - SIMILARITY: Contains 12 VWFA domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X52022; CAA36267.1; -;  
CC EMBL: X06196; CAA29557.1; -;  
CC EMBL: M20778; -; NOT ANNOTATED\_CDS.  
CC EMBL: M27449; AAA52057.1; -;  
CC EMBL: S49432; AAA24261.1; -;  
CC FIR: A59140; CGHU3A.  
CC -----

DR PDB; 1KNT; 01-NOV-94.  
DR PDB; 2KNT; 15-MAY-97.  
DR PDB; 1KUN; 12-NOV-97.  
DR PDB; 1KTH; 28-AUG-02.  
DR Genew; HGNC:2213; COL6A3.  
DR MIM; 120250; -.  
DR MIM; 158810; -.  
DR MIM; 254090; -.  
DR GO; 0005589; C:collagen type VI; TAS.  
DR GO; 0007517; P:muscle development; TAS.  
DR InterPro; IPR008161; Clg\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR003961; FN III.  
DR InterPro; IPR02223; Kunitz\_BPTI.  
DR InterPro; IPR02035; VWF\_A.  
DR Pfam; PF01391; Collagen; 5.  
DR Pfam; PF00014; Kunitz\_BPTI; 1.  
DR Pfam; PF00092; vwa; 11.  
DR PRINTS; PR00759; BASICPTASE.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR PRODOM; PD000007; Clg\_helix; 2.  
DR PRODOM; PD000222; Kunitz\_BPTI; 1.  
DR SMART; SM00060; FN3; 1.  
DR SMART; SM00131; KU; 1.  
DR SMART; SM00327; VWA; 12.  
DR PROSITE; PS00280; BPTI\_KUNITZ\_1; 1.  
DR PROSITE; PS02079; BPTI\_KUNITZ\_2; 1.  
DR PROSITE; PS02034; VWFA; 12.  
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Glycoprotein; Cell adhesion; Serine protease inhibitor; Collagen;  
KW Signal; 3D-structure; Disease mutation; Polymorphism;  
KW Alternative splicing.  
FT SIGNAL 1 25 POTENTIAL.  
FT CHAIN 26 3176 COLLAGEN ALPHA 3(VI) CHAIN.  
FT DOMAIN 26 2038 NONHELICAL REGION.  
FT DOMAIN 2039 2375 TRIPLE-HELICAL REGION.  
FT DOMAIN 2376 3176 NONHELICAL REGION.  
FT DOMAIN 39 213 VWFA 1.  
FT DOMAIN 242 419 VWFA 2.  
FT DOMAIN 445 620 VWFA 3.  
FT DOMAIN 639 816 VWFA 4.  
FT DOMAIN 837 1009 VWFA 5.  
FT DOMAIN 1029 1205 VWFA 6.  
FT DOMAIN 1233 1404 VWFA 7.  
FT DOMAIN 1436 1609 VWFA 8.  
FT DOMAIN 1639 1812 VWFA 9.  
FT DOMAIN 1838 2024 VWFA 10.  
FT DOMAIN 2402 2581 VWFA 11.  
FT DOMAIN 2619 2815 VWFA 12.  
FT DOMAIN 2987 3076 FIBRONECTIN TYPE-III.  
FT DOMAIN 3107 3176 BPTI/KUNITZ INHIBITOR.  
FT SITE 2040 2042 CELL ATTACHMENT SITE.  
FT SITE 2136 2138 CELL ATTACHMENT SITE.  
FT SITE 2148 2150 CELL ATTACHMENT SITE.  
FT SITE 2154 2156 CELL ATTACHMENT SITE.  
FT SITE 2370 2372 CELL ATTACHMENT SITE.  
FT ACT SITE 3121 3122 REACTIVE BOND.  
FT DISULFID 3111 3161  
FT DISULFID 3120 3144  
FT DISULFID 3136 3157  
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 40.5%; Score 42.5; DB 1; Length 3176;  
Best Local Similarity 47.1%; Pred. No. 4.5e+02;  
Matches 8; Conservative 4; Mismatches 4; Indels 1; Gaps 1;  
QY 2 PAKGMSPPGFIVGEEV 18  
Db 2232 PAKGAPPG-LIGEGGI 2247.

Search completed: March 22, 2004, 06:53:02  
Job time : 0.767093 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:39:53 ; Search time 4.02446 Seconds  
(without alignments)  
1568.003 Million cell updates/sec

Title: US-09-662-293-13  
Perfect score: 105  
Sequence: 1 DPAKMGPPGFIVGEEGVLS 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 25:  
1: sp\_archaea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_mbc:  
8: sp\_organelle:  
9: sp\_phage:  
10: sp\_plant:  
11: sp\_rodent:  
12: sp\_virus:  
13: sp\_vertebrate:  
14: sp\_unclassified:  
15: sp\_virus:  
16: sp\_bacteriap:  
17: sp\_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	99	94.3	555	Q3U6R7	Q3U6R7 dermatophag
2	50	47.6	408	Q8MR29	Q8MR29 drosophila
3	50	47.6	408	Q8W5E2	Q8W5E2 drosophila
4	48	45.7	118	Q7Z360	Q7Z360 homo sapien
5	47	44.8	201	Q825J2	Q825J2 nitrosomona
6	47	44.8	573	Q9X141	Q9X141 arabidopsis
7	47	44.8	584	Q8RXP2	Q8RXP2 arabidopsis
8	47	44.8	585	Q8H1F3	Q8H1F3 arabidopsis
9	47	44.8	880	Q8E7U2	Q8E7U2 streptococc
10	47	44.8	880	Q8E3D5	Q8E3D5 streptococc
11	46.5	44.3	198	Q7X4K8	Q7X4K8 synechococc
12	46	43.8	144	Q8RJY9	Q8RJY9 stigmatella
13	46	43.8	433	Q9A5K3	Q9A5K3 caulobacter
14	46	43.8	455	Q8HB49	Q8HB49 oryza sativ
15	46	43.8	662	Q974C3	Q974C3 sulfolobus
16	46	43.8	3075	Q8AW10	Q8AW10 brachydanio

17	46	43.8	4641	4	Q75592	Q75592 homo sapien
18	46	43.8	4708	11	Q7TPH6	Q7TPH6 mus musculu
19	45	42.9	79	16	Q98JT1	Q98JT1 rhizobium l
20	45	42.9	115	12	Q39919	Q39919 human cytom
21	45	42.9	174	16	Q8A3Y9	Q8A3Y9 bacteroides
22	45	42.9	283	16	Q8ALP1	Q8ALP1 bacteroides
23	45	42.9	303	16	Q8NMD1	Q8NMD1 corynebacte
24	45	42.9	351	16	Q8PPL3	Q8PPL3 xanthomonas
25	45	42.9	353	5	Q8MZL0	Q8MZL0 drosophila
26	45	42.9	402	11	Q91Z98	Q91Z98 mus musculu
27	45	42.9	402	11	Q8VH43	Q8VH43 mus musculu
28	45	42.9	424	16	Q89RH4	Q89RH4 bradyrhizob
29	45	42.9	494	2	Q9ZNU2	Q9ZNU2 aeromonas s
30	45	42.9	626	16	Q8D977	Q8D977 vibrio vuln
31	45	42.9	634	16	Q87P74	Q87P74 vibrio para
32	45	42.9	635	16	Q92KX0	Q92KX0 rhizobium m
33	45	42.9	785	5	Q7YYE6	Q7YYE6 cryptospori
34	44.5	42.4	6298	11	Q8VHN7	Q8VHN7 mus musculu
35	44	41.9	111	17	Q9YAE9	Q9YAE9 aeropyrum p
36	44	41.9	159	17	Q8PXB0	Q8PXB0 methanogarc
37	44	41.9	256	5	Q9NCB5	Q9NCB5 tribolium c
38	44	41.9	306	16	Q87M92	Q87M92 vibrio para
39	44	41.9	342	4	Q13074	Q13074 homo sapien
40	44	41.9	375	3	Q873B0	Q873B0 neurospora
41	44	41.9	403	17	Q8ZXI6	Q8ZXI6 pyrobaculum
42	44	41.9	405	17	Q8TZU1	Q8TZU1 pyrococcus
43	44	41.9	408	17	Q97ZM0	Q97ZM0 sulfolobus
44	44	41.9	410	17	Q9YAE1	Q9YAE1 aeropyrum p
45	44	41.9	444	16	Q98IL7	Q98IL7 rhizobium l

ALIGNMENTS

RESULT 1

Q9U6R7  
ID Q9U6R7 PRELIMINARY; PRT; 555 AA.  
AC Q9U6R7;  
DT 01-MAY-2000 (TREMREL. 13, Created)  
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)  
DT 01-OCT-2003 (TREMREL. 25, Last annotation update)  
DE 98kDa HDM allergen.  
OS Dermatophagoides farinae (House-dust mite).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
OC Acariformes; Sarcoptiformes; Astigmata; Psoroptidia; Analgoidea;  
OC Pyroglyphidae; Dermatophagoides.  
CX NCBI\_TaxID=6954;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Weber E.R., Hunter S., Stedman K., McCall C.;  
RT "Cloning and Characterization of a 98 kDa Allergen from  
RT Dermatophagoides farinae."  
RL Submitted (JUG-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF178772; AAD5672.1; ...  
DR GO; GO:0005576; C:extracellular; IEA.  
DR GO; GO:0008061; F:chitin binding; IEA.  
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR GO; GO:0006030; P:chitin metabolism; IEA.  
DR InterPro; IPR002557; Chitin\_bind\_perA.  
DR InterPro; IPR01223; Glyco\_Hydro\_18.  
DR InterPro; IPR01579; Glyco\_Hydro\_18A.  
DR Pfam; PF00704; Glyco\_hydro\_18; 1.  
DR ProDom; PD000471; Glyco\_hydro\_18; 1.  
DR SMART; SM00494; ChnBD2; 1.  
DR SMART; SM00636; Glyco\_18; 1.  
DR PROSITE; PS01095; CHITINASE\_18; 1.  
KW Glycosidase; Hydrolase.  
SQ SEQUENCE 555 AA; 63238 MW; 0E4564A1A459B30B CRC64;

Query Match 94.3%; Score 99; DB 5; Length 555;  
Best Local Similarity 95.0%; Pred. No. 3.7e-07;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;



```

QY 1 DPAKGMSPPGFIVGEGVLS 20
    |||||
Db 298 DPAKGMSPPGFIVGEGVLS 317

RESULT 2
Q8MRZ9 PRELIMINARY; PRT; 408 AA.
AC Q8MRZ9;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE RH61522P.
GN EG:115C2.8 OR CG13367.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dreesen D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Pargass V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AV119176; AAM51036.1; -.
DR FlyBase; FBgn0025634; EG:115C2.8.
SQ SEQUENCE 408 AA; 43953 MW; 0FD51AE3C6303960 CRC64;

Query Match 47.6%; Score 50; DB 5; Length 408;
Best Local Similarity 40.0%; Pred. No. 16;
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 DPAKGMSPPGFIVGEGVLS 20
    |||||
Db 327 DPAKGMSPPGFIVGEGVLS 346

RESULT 3
Q9W5E2 PRELIMINARY; PRT; 408 AA.
AC Q9W5E2;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE EG:115C2.8 protein.
GN EG:115C2.8 OR CG13367.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Salles C., Valenti P., Darlamitsou A., Henderson N., Campbell L.,
RA Glover D.;
RL "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Benos P.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003418; AAF45535.1; -.
DR EMBL; AL031581; CAA20887.1; -.
DR PIR; T13387; T13387.
DR FlyBase; FBgn0025634; EG:115C2.8.
SQ SEQUENCE 408 AA; 43939 MW; 8B980E52E0886F4F CRC64;

Query Match 47.6%; Score 50; DB 5; Length 408;
Best Local Similarity 40.0%; Pred. No. 16;
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 DPAKGMSPPGFIVGEGVLS 20
    |||||
Db 327 DPAKGMSPPGFIVGEGVLS 346

RESULT 4
Q7Z360 PRELIMINARY; PRT; 118 AA.
AC Q7Z360;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hypothetical protein DKFZp686I01145.
GN DKFZp686I01145.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Human colon endothel primary cell culture;
RA Bloeker H., Bocher M., Mewes H.W., Weil B., Amid C., Osanger A.,
RA Fobo G., Han M., Wiemann S.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BX538104; CAD98017.1; -.
KW Hypothetical protein.
SQ SEQUENCE 118 AA; 12808 MW; 0310EF37E1F0E92C CRC64;

Query Match 45.7%; Score 48; DB 4; Length 118;
Best Local Similarity 43.8%; Pred. No. 9;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 PAKGMSPPGFIVGEG 17
    |||||
Db 76 PSEGLCPGHHISEDG 91

RESULT 5
Q82SJ2 PRELIMINARY; PRT; 201 AA.
AC Q82SJ2;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN NE2330.
OS Nitrosomonas europaea.
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC Nitrosomonadaceae; Nitrosomonas.
OX NCBI_TaxID=915;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19718 / IFO 14298;
RX MEDLINE=22586410; PubMed=12700255;
RA Chain P., Lanerkin J.E., Larimer F.W., Regala W., Lao V., Land M.,
RA Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,
RA Arciero D.M., Hommes N.G., Whittaker M.M., Arp D.J.;
RT "Complete genome sequence of the ammonia-oxidizing bacterium and
RT obligate chemolithoautotroph Nitrosomonas europaea.";
RL J. Bacteriol. 185:2759-2773(2003).
DR EMBL; BX321864; CAD86242.1; -.
DR InterPro; IPR005572; RseA_N.
DR Pfam; PF03872; RseA_N; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 201 AA; 22079 MW; C5364D2B5BCD1A45 CRC64;

Query Match 44.8%; Score 47; DB 16; Length 201;
Best Local Similarity 47.4%; Pred. No. 23;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
```

Query Match 44.8%; Score 47; DB 10; Length 584;  
Best Local Similarity 47.1%; Pred. No. 73;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 PAKGMSPPGFIVGEGV 18  
DB 477 PTFGMAYPGVQSGEDGV 493

RESULT 8

Q8H1F3 PRELIMINARY; PRT; 585 AA.

AC Q8H1F3  
DT 01-MAR-2003 (TREMELrel. 23, Created)  
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
DE Hypothetical protein.  
GN At1G15280  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,  
RA Quach H.L., Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,  
RA Yu G., Yuan S., Chen H., Cheuk R., Jones T., Kim C.J., Nguyen M.,  
RA Palm C.J., Shinn P., Southwick A., Tripp M.G., Wu T., Davis R.W.,  
RA Ecker J.R., Theologis A.;  
RT "Arabidopsis Open Reading Frame (ORF) Clones."  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
KW Hypothetical protein.  
SQ SEQUENCE 585 AA; 63435 MW; 5E0FF3FBD185A15 CRC64;

Query Match 44.8%; Score 47; DB 10; Length 585;  
Best Local Similarity 47.1%; Pred. No. 73;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 PAKGMSPPGFIVGEGV 18  
DB 478 PTFGMAYPGVQSGEDGV 494

RESULT 9

Q8E7U2 PRELIMINARY; PRT; 880 AA.

AC Q8E7U2  
DT 01-MAR-2003 (TREMELrel. 23, Created)  
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN At1G15280  
OS Streptococcus agalactiae (serotype III).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus  
OX NCBI\_TaxID=216495;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,  
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.;  
RT "Arabidopsis Full Length cDNA Clones."  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
RW Hypothetical protein.  
SQ SEQUENCE 584 AA; 63284 MW; 5DD94B2B058B343 CRC64;

Query Match 44.8%; Score 47; DB 10; Length 573;  
Best Local Similarity 47.1%; Pred. No. 71;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 PAKGMSPPGFIVGEGV 18  
DB 477 PTFGMAYPGVQSGEDGV 493

RESULT 7

Q8RXP2 PRELIMINARY; PRT; 584 AA.

AC Q8RXP2  
DT 01-JUN-2002 (TREMELrel. 21, Created)  
DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN At1G15280  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,  
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.;  
RT "Arabidopsis Full Length cDNA Clones."  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
RW Hypothetical protein.  
SQ SEQUENCE 584 AA; 63284 MW; 5DD94B2B058B343 CRC64;

```

DR InterPro; IPR002086; Aldehyde_dehydr.
DR InterPro; IPR001670; Fe-ADH.
DR Pfam; PF00171; aldedh; 1.
DR Pfam; PF00465; Fe-ADH; 1.
DR PROSITE; PS00913; ADH_IRON_1; 1.
DR PROSITE; PS00060; ADH_IRON_2; 1.
DR PROSITE; PS00060; ADH_IRON_2; 1.
DR HypoNet; HYP00000; Complete_proteome.
DR KW
DR SEQUENCE 880 AA; 97005 MW; FC26D6B9F7FF384C CRC64;

Query Match 44.8%; Score 47; DB 16; Length 880;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 DPAKMGSPPGFIVGEGV 18
||| : |||| : |
DB 647 DPALVMTVPGFIAADTCM 664

RESULT 10
Q8E2D5 PRELIMINARY; PRT; 880 AA.
AC Q8E2D5
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Aldehyde-alcohol dehydrogenase.
GN ADHE OR SAG0053.
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
SEQUENCE FROM N.A.
RA STRAIN=2603 V/R / Serotype V;
RX MEDLINE=22222988; PubMed=12200547;
RA Tettelin H., Maignan V., Chelstewicz M.J., Eisen J.A., Peterson S.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Malone D.,
RA Rinaudo P., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
DR EMBL; AE014194; AA98961.1; -.
DR TIGR; SAG0053; -.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016431; F:oxidoreductase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR InterPro; IPR001670; Fe-ADH.
DR Pfam; PF00171; aldedh; 1.
DR Pfam; PF00465; Fe-ADH; 1.
DR PROSITE; PS00913; ADH_IRON_1; 1.
DR PROSITE; PS00060; ADH_IRON_2; 1.
DR KW
DR Complete proteome.
DR SEQUENCE 880 AA; 97005 MW; FC26D6B9F7FF384C CRC64;

Query Match 44.8%; Score 47; DB 16; Length 880;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 DPAKMGSPPGFIVGEGV 18
||| : |||| : |
DB 647 DPALVMTVPGFIAADTCM 664

RESULT 11
Q7X4K8 PRELIMINARY; PRT 198 AA.
ID Q7X4K8

```

```

QY      2  PAKGMSPPGFVIGEEGV 18
      |:|:|:|:|:|:|:|:|:|:|
DB      391  PSRGMTPTGYCVTAAGV 407

RESULT 15
Q974C3  PRELIMINARY;      PRT;      662 AA.
Q974C3;
DT      01-DEC-2001 (TrEMBLrel. 19, Created)
DT      01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Putative acetyl-CoA synthetase.
GN      ST0730.
OS      Sulfolobus tokodaii.
OC      Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC      Sulfolobus.
OX      NCBI_TaxID=111955;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN:JCM 10545 / 7;
RX      MEDLINE=21456156; PubMed=11572479;
RA      Kawarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA      Sekine M., Baba S.-I., Ankaï A., Kosugi H., Hosoyama A., Fukui S.,
RA      Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA      Yoshizawa T., Tanaka T., Kudon Y., Yanazaki J., Kushida N., Oguchi A.,
RA      Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA      Oshima T., Kikuchi H.;
RT      "Complete genome sequence of an aerobic thermoacidophilic
RT      Crenarchaeon, Sulfolobus tokodaii strain7.";
RL      DNA Res. 8:123-140(2001).
RL      EMBL; AF000983; BAB65737.1; -.
DR      GO; GO:0003824; P:catalytic activity; IEA.
DR      GO; GO:0008152; P:metabolism; IEA.
DR      InterPro; IPR000873; AMP-bind.
DR      Pfam; PF00501; AMP-binding; 1.
DR      PRINTS; PR00154; AMPBINDING.
DR      PROSITE; PS00455; AMP BINDING; 1.
KW      Hypothetical protein, Complete proteome.
SQ      SEQUENCE 662 AA; 74544 MW; FC698FB9AEA2C107 CRC64;

Query Match      43.8%; Score 46; DB 17; Length 662;
Best Local Similarity 44.4%; Pred. NO. 1.2e-02;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY      1  DPAKGMSPGFVIGEEGV 18
      |:|:|:|:|:|:|:|:|:|:|
DB      561  DEVKGEVPAFVILKQGV 578

Search completed: March 22, 2004, 06:59:22
Job time : 4.02446 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2004, 06:30:23 ; Search time 6.08116 Seconds  
(without alignments)  
929.256 Million cell updates/sec

Title: US-09-662-293-13

Perfect score: 105

Sequence: 1 DPAKGMSPGPIVGBEGLVS 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 29Jan04:\*

1: Geneseqp1980s:\*

2: Geneseqp1990s:\*

3: Geneseqp2000s:\*

4: Geneseqp2001s:\*

5: Geneseqp2002s:\*

6: Geneseqp2003as:\*

7: Geneseqp2003bs:\*

8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	105	100.0	20	AA52513	House dus
2	105	100.0	20	AA52522	House dus
3	105	100.0	20	AAU96317	Der HWM-m
4	105	100.0	20	AAU96326	Der HWM-m
5	101	96.2	490	AA52535	D. pteron
6	101	96.2	490	AAU96339	Der HWM-m
7	101	96.2	509	AA52533	D. pteron
8	101	96.2	509	AAU96337	Der HWM-m
9	101	96.2	509	AAU96338	Der HWM-m
10	99	94.3	536	AA52525	House dus
11	99	94.3	536	AAU96329	Der HWM-m
12	99	94.3	555	AA52523	House dus
13	99	94.3	555	AAU96327	Der HWM-m
14	99	94.3	555	AAU96328	Der HWM-m
15	50	47.6	488	AB68081	Drosophil
16	48	45.7	57	AA58145	Lung canc
17	48	45.7	117	ABG71824	Wild type
18	48	45.7	130	AAU99897	TNF relat
19	48	45.7	350	AAU00934	Human DR5
20	48	45.7	411	AAW76827	Human TR6
21	48	45.7	411	AAW79083	Human dea
22	48	45.7	411	AAW79261	Tumour ne
23	48	45.7	411	AAW88410	Human Apo
24	48	45.7	411	AAW83321	Human Apo
25	48	45.7	411	AAW93576	Human hAP

#### ALIGNMENTS

##### RESULT 1

AA52513

ID AA52513 standard; peptide; 20 AA.

XX AC AA52513;

XX AC

DT 22-FEB-2000 (first entry)

DE House dust mite allergen protein (map) A/B fragment map(3).

XX KW Mite allergen protein; map; high molecular weight; HMW-map; allergy;

XX KW house dust mite; IgE; immunoglobulin E; allergen; mapA; mapB;

XX KW hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;

XX KW canine; veterinary; antibody; vaccine; immunisation.

XX OS Dermatophagoides farinae.

XX PN WO9954349-A2.

XX PD 28-OCT-1999.

XX PF 16-APR-1999; 99WO-US008524.

PR 17-APR-1998; 98US-00062013.

PR 13-MAY-1998; 98US-0085295P.

PR 02-SEP-1998; 98US-0098909P.

PA (HESK-) HESKA CORP.

XX Mccall CA, Hunter SW, Weber ER;

XX WPI; 2000-052700/04.

PT Novel high molecular weight Dermatophagoides nucleic acid polypeptides

PT used to modify an animals' hypersensitivity to mite allergens.

XX Claim 3; Page 69; 154pp; English.

XX Sequences AA52510-Y52522 represent proteolytic fragments of

CC Dermatophagoides farinae high molecular weight mite allergen protein (HMW

CC -map) composition. The HMW-map composition was isolated from a D. farinae

CC homogenate by gel filtration, with each fraction being analysed for the

CC presence of proteins that bound to IgE present in mite-allergic dog

CC antisera. The HMW-map composition comprises mapA (a 109 kD protein) and

CC mapB (98 kD). Mite allergenic proteins and peptides, and nucleic acids

CC encoding them, may be used in therapeutic compositions to modify an

CC animal's hypersensitivity reaction to mite allergens. Animals that may be

CC treated include mammals and birds, especially felines, canines, equines,

XX

```

CC allergens from a mixture of proteins
SQ Sequence 20 AA;

Query Match 100.0%; Score 105; DB 3; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPAKGMSPPGFIVGEEGVLS 20
DB |||||
1 DPAKGMSPPGFIVGEEGVLS 20

RESULT 3
AAU96317
ID AAU96317 standard; peptide; 20 AA.
XX AC AAU96317;
XX AC
XX DT 15-JUL-2002 (first entry)
XX DE Der HMW-map polypeptide #4.
XX DE Der HMW-map; American house dust mite; antiallergic; mite; IgE;
KW mite allergenic protein; immunoglobulin E; hypersensitivity;
KW immunocomplex formation.
XX KW
XX KW Dermatophagoides farinae.
XX OS WO200222807-A2.
XX EN
XX FN 21-MAR-2002.
XX FD
XX FF 14-SEP-2001; 2001WO-US028730.
XX FF
XX PR 14-SEP-2000; 2000US-00662293.
XX PR
XX XX (HESK-) HESKA CORP.
XX XX McCall CA, Hunter SW, Weber ER;
XX XX WPI; 2002-351888/38.
XX DR
XX PT New mite allergenic protein isolated from Dermatophagoides, designated
XX PT Der HMW-map protein, useful as a vaccine for treating mite allergy.
XX XX
XX PS Claim 12; Page 70; 16app; English.
XX CC The invention relates to an isolated mite allergenic protein of
XX CC Dermatophagoides, designated Der HMW-map protein, and its related nucleic
XX CC acid. The Der HMW-map protein is useful for eliciting an immune response
XX CC against Der HMW-map protein. The protein or a reagent comprising a non-
XX CC proteinaceous epitope is useful for identifying an animal (e.g., dog,
XX CC cat) susceptible to or having an allergic response to a mite. A
XX CC therapeutic composition is useful for desensitizing a host animal to an
XX CC allergic response to a mite. The DNA and protein can be used in the
XX CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition
XX CC of immunoglobulin (Ig)E or Der HMW-map protein activity associated with a
XX CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting
XX CC binding of proteins to IgE, to prevent immunocomplex formation, thus
XX CC reducing hypersensitivity responses to mite allergens, and as vaccines
XX CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342
XX CC represent Der HMW-map polypeptides of the invention
SQ Sequence 20 AA;

Query Match 100.0%; Score 105; DB 5; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPAKGMSPPGFIVGEEGVLS 20
DB |||||
1 DPAKGMSPPGFIVGEEGVLS 20

```

DE	D. pteronyssius 98 kD mite allergen protein (map) PDerp98-490.
XX	
KW	Mite allergen protein; map; high molecular weight; HMW-map; allergy;
KW	house dust mite; IgE; immunoglobulin E; allergen; mapB;
KW	hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;
KW	canine; veterinary; antibody; vaccine; immunisation.
XX	
OS	Dermatophagoideis pteronyssinus.
XX	
XX	Key Location/Qualifiers
FT	Modified-site 115..117
FT	/note="Asn is N-glycosylated"
FT	Modified-site 240..242
FT	/note="Asn is N-glycosylated"
XX	
PN	WO9954349-A2.
XX	
XX	28-OCT-1999.
XX	
PF	16-APR-1999; 99WO-US008524.
XX	
PR	17-APR-1998; 98US-00062013.
PR	13-MAY-1998; 98US-0085295P.
PR	02-SEP-1998; 98US-0098909P.
XX	
XX	(HESK-) HESKA CORP.
PA	
XX	Mccall CA, Hunter SW, Weber BR;
PI	
DR	WPI; 2000-052700/04.
DR	N-PSDB; AA238589, AA238590.
XX	
PT	Novel high molecular weight Dermatophagoideis nucleic acid polypeptides
PT	used to modify an animals' hypersensitivity to mite allergens.
XX	
PS	Claim 3; Page 147-149; 154pp; English.
XX	
CC	This sequence represents Dermatophagoideis pteronyssius mite allergen
CC	protein (map) PDerp98-490, the mature form of pDerp98-509. PDerp98-490
CC	has a molecular weight of 98 kD, comprising 490 amino acids, and has a
CC	high degree of homology with the D. farinae mature 98 kD allergen, mapB
CC	(AA252525). Nucleic acid molecules encoding Pderp98-490 were isolated
CC	from a D. pteronyssius cDNA library by hybridisation with a probe
CC	encoding the D. farinae high molecular weight map (HMW-map) composition.
CC	Mite allergenic proteins and peptides, and nucleic acids encoding them,
CC	may be used in therapeutic compositions to modify an animal's
CC	hypersensitivity reaction to mite allergens. Animals that may be treated
CC	include mammals and birds, especially felines, canines, equines, humans,
CC	other pets, and work or domestic animals. The proteins or fragments may
CC	also be used to diagnose allergies via a skin test. The proteins and
CC	peptides can also be used to raise antibodies, which have a variety of
CC	potential uses. For example, they can be used as vaccines to passively
CC	immunise animals against dust mite hypersensitivity, as positive controls
CC	in test kits and as tools to recover desired dust mite allergens from a
CC	mixture of proteins. (Updated on 06-AUG-2003 to correct OS field.)
XX	
SQ	Sequence 490 AA;
	Query Match 96.2%; Score 101; DB 3; Length 490;
	Best Local Similarity 95.0%; Pred. No. 4.1e-07;
	Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY	1 DPAKGNPPGFIIVEGVLS 20
Db	279 DPAKGNPPGFIIVEGVLS 298
RESULT 6	
AAU96339	
ID AAU96339	standard; protein; 490 AA.
XX AC	
XX AAU96339;	
XX	

DT 15-JUL-2002 (first entry)  
XX Der HMW-map polypeptide #26.  
DE  
XX Der HMW-map; American house dust mite; antiallergic; mite; IgE;  
XX mite allergenic protein; immunoglobulin E; hypersensitivity;  
KW immunocomplex formation.  
KW  
XX Dermatophagoides farinae.  
OS  
XX WO200222807-A2.  
PN  
XX 21-MAR-2002.  
XX  
XX 14-SEP-2001; 2001WO-US028730.  
PF  
XX 14-SEP-2000; 2000US-00662293.  
PR  
XX (HESK-) HESKA CORP.  
XX  
XX McCall CA, Hunter SW, Weber ER;  
PI  
XX WPI; 2002-351888/38.  
DR  
XX N-PSDB; ABK69585.  
DR  
XX New mite allergenic protein isolated from Dermatophagoides, designated  
PT Der HMW-map protein, useful as a vaccine for treating mite allergy.  
PT  
XX Claim 12; Page 144-146; 161pp; English.  
PS  
XX The invention relates to an isolated mite allergenic protein of  
XX Dermatophagoides, designated Der HMW-map protein, and its related nucleic  
CC acid. The Der HMW-map protein is useful for eliciting an immune response  
CC against Der HMW-map protein. The protein or a reagent comprising a non-  
CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
CC cat) susceptible to or having an allergic response to a mite.  
CC A  
CC therapeutic composition is useful for desensitising a host animal to an  
CC allergic response to a mite. The DNA and protein can be used in the  
CC detection of anti-Der HMW-map antibodies in animal fluids, and inhibition  
CC of immunoglobulin (Ig) E or Der HMW-map protein activity associated with a  
CC disease. Antibodies that bind to Der HMW-map are useful for inhibiting  
CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
CC reducing hypersensitivity responses to mite allergens, and as vaccines  
CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
CC represent Der HMW-map polypeptides of the invention  
XX  
XX Sequence 490 AA;  
SQ  
Query Match 96.2%; Score 101; DB 5; Length 490;  
Best Local Similarity 95.0%; Pred. No. 4.1e-07;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 DPAKGMSPPGFTVGEGLVLS 20  
DB 279 DPAKGMSPPGFTVGEGLVLS 298  
RESULT 7  
AAU52533  
ID AAU52533 standard; protein; 509 AA.  
XX  
XX AAU52533;  
AC  
DT 06-AUG-2003 (revised)  
DT 22-FEB-2000 (first entry)  
XX  
XX D. pteronyssius 98 kD mite allergen protein (map) PDerp98-509.  
XX  
XX Mite allergen protein; map; high molecular weight; HMW-map; allergy;  
KW house dust mite; IgE; immunoglobulin E; allergen; mapB;  
KW hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
KW canine; veterinary; antibody; vaccine; immunisation.  
KW  
KW

OS Dermatophagoides pteronyssinus.  
XX  
XX Key Location/Qualifiers  
PH Peptide 1..19  
FT /note= "Signal peptide"  
FT Protein 20..509  
FT /note= "Mature PDerp98-509"  
XX  
XX WO9954349-A2.  
PN  
XX 28-OCT-1999.  
PD  
XX 16-APR-1999; 99WO-US008524.  
PF  
XX 17-APR-1998; 98US-00062013.  
PR 13-MAY-1998; 98US-0085295F.  
PR 02-SEP-1998; 98US-0098909F.  
XX  
XX (HESK-) HESKA CORP.  
PA  
XX McCall CA, Hunter SW, Weber ER;  
PI  
XX WPI; 2000-052700/04.  
DR  
XX N-PSDB; AAZ38585, AAZ38586, AAZ38587, AAZ38588.  
DR  
XX Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
PT used to modify an animals' hypersensitivity to mite allergens.  
PT  
XX Claim 3; Page 134-136; 154pp; English.  
PS  
XX This sequence represents Dermatophagoides pteronyssius mite allergen  
XX protein (map) PDerp98-509. PDerp98-509 has a molecular weight of 98 kD, D.  
CC comprising 509 amino acids, and has a high degree of homology with the  
CC farinae 98 kD allergen, mapB (AAV52523). Nucleic acid molecules encoding  
CC PDerp98-509 were isolated from a D. pteronyssius cDNA library by  
CC hybridisation with a probe encoding the D. farinae high molecular weight  
CC map (HMW-map) composition. Mite allergenic proteins and peptides, and  
CC nucleic acids encoding them, may be used in therapeutic compositions to  
CC modify an animal's hypersensitivity reaction to mite allergens. Animals  
CC that may be treated include mammals and birds, especially felines,  
CC canines, equines, humans, other pets, and work or domestic animals. The  
CC proteins or fragments may also be used to diagnose allergies via a skin  
CC test. The proteins and peptides can also be used to raise antibodies, as  
CC which have a variety of potential uses. For example, they can be used as  
CC vaccines to passively immunise animals against dust mite  
CC hypersensitivity, as positive controls in test kits and as tools to  
CC recover desired dust mite allergens from a mixture of proteins. (Updated  
CC on 06-AUG-2003 to correct OS field.)  
XX  
XX Sequence 509 AA;  
SQ  
Query Match 96.2%; Score 101; DB 3; Length 509;  
Best Local Similarity 95.0%; Pred. No. 4.2e-07;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 DPAKGMSPPGFTVGEGLVLS 20  
DB 298 DPAKGMSPPGFTVGEGLVLS 317  
RESULT 8  
AAU96337  
ID AAU96337 standard; protein; 509 AA.  
XX  
XX AAU96337;  
AC  
XX 15-JUL-2002 (first entry)  
DT  
XX Der HMW-map polypeptide #24.  
DE  
XX Der HMW-map; American house dust mite; antiallergic; mite; IgE;  
KW mite allergenic protein; immunoglobulin E; hypersensitivity;  
KW immunocomplex formation.  
KW



```

XX OS Dermatophagoides farinae.
XX PN WO200222807-A2.
XX PD 21-MAR-2002.
XX XX
XX PF 14-SEP-2001; 2001WO-US028730.
XX PR 14-SEP-2000; 2000US-00662293.
XX PA (HESK-) HESKA CORP.
XX PI Mccall CA, Hunter SW, Weber ER;
XX DR WPI; 2002-351888/38.
XX DR N-PSDB; ABK69581.
XX XX
XX PT New mite allergenic protein isolated from Dermatophagoides, designated
XX PT Der HW-map protein, useful as a vaccine for treating mite allergy.
XX PS Claim 12; Page 134-136; 161pp; English.
XX CC
XX CC The invention relates to an isolated mite allergenic protein of
XX CC Dermatophagoides, designated Der HW-map protein, and its related nucleic
XX CC acid. The Der HW-map protein is useful for eliciting an immune response
XX CC against Der HW-map protein. The protein or a reagent comprising a non-
XX CC proteinaceous epitope is useful for identifying an animal (e.g., dog,
XX CC cat) susceptible to or having an allergic response to a mite. A
XX CC therapeutic composition is useful for desensitising a host animal to an
XX CC allergic response to a mite. The DNA and protein can be used in the
XX CC detection of anti-Der HW-map antibodies in animal fluids, and inhibition
XX CC of immunoglobulin (Ig)E or Der HW-map protein activity associated with a
XX CC disease. Antibodies that bind to Der HW-map are useful for inhibiting
XX CC binding of proteins to IgE, to prevent immunocomplex formation, thus
XX CC reducing hypersensitivity responses to mite allergens, and as vaccines
XX CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342
XX CC represent Der HW-map polypeptides of the invention
XX SQ Sequence 509 AA;
XX
Query Match 96.2%; Score 101; DB 5; Length 509;
Best Local Similarity 95.0%; Pred. No. 4.2e-07;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DPAKGMSPPGFIVGEGVLS 20
Db ||||| ||||| ||||| ||||| |||||
298 DPAKGMSPPGFIVGEGVLS 317

RESULT 9
AAU96338
ID AAU96338 standard; protein; 509 AA.
XX AC AAU96338;
XX DT 15-JUL-2002 (first entry)
XX DE Der HW-map polypeptide #25.
XX KW Der HW-map; American house dust mite; antiallergic; mite; IgE;
XX KW mite allergenic protein; immunoglobulin E; hypersensitivity;
XX KW immunocomplex formation.
XX OS Dermatophagoides farinae.
XX PN WO200222807-A2.
XX PD 21-MAR-2002.
XX PF 14-SEP-2001; 2001WO-US028730.
XX PR 14-SEP-2000; 2000US-00662293.

```

```

XX PA (HESK-) HESKA CORP.
XX PI Mccall CA, Hunter SW, Weber ER;
XX XX
XX WPI; 2002-351888/38.
XX DR N-PSDB; ABK69583.
XX PT New mite allergenic protein isolated from Dermatophagoides, designated
XX PT Der HW-map protein, useful as a vaccine for treating mite allergy.
XX PS Claim 12; Page 139-141; 161pp; English.
XX CC
XX CC The invention relates to an isolated mite allergenic protein of
XX CC Dermatophagoides, designated Der HW-map protein, and its related nucleic
XX CC acid. The Der HW-map protein is useful for eliciting an immune response
XX CC against Der HW-map protein. The protein or a reagent comprising a non-
XX CC proteinaceous epitope is useful for identifying an animal (e.g., dog,
XX CC cat) susceptible to or having an allergic response to a mite. A
XX CC therapeutic composition is useful for desensitising a host animal to an
XX CC allergic response to a mite. The DNA and protein can be used in the
XX CC detection of anti-Der HW-map antibodies in animal fluids, and inhibition
XX CC of immunoglobulin (Ig)E or Der HW-map protein activity associated with a
XX CC disease. Antibodies that bind to Der HW-map are useful for inhibiting
XX CC binding of proteins to IgE, to prevent immunocomplex formation, thus
XX CC reducing hypersensitivity responses to mite allergens, and as vaccines
XX CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342
XX CC represent Der HW-map polypeptides of the invention
XX SQ Sequence 509 AA;
XX
Query Match 96.2%; Score 101; DB 5; Length 509;
Best Local Similarity 95.0%; Pred. No. 4.2e-07;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DPAKGMSPPGFIVGEGVLS 20
Db ||||| ||||| ||||| ||||| |||||
298 DPAKGMSPPGFIVGEGVLS 317

RESULT 10
AAU96338
ID AAU96338 standard; protein; 536 AA.
XX AC AAU96338;
XX DT 22-FEB-2000 (first entry)
XX DE House dust mite (D. farinae) mite allergen protein (map) PDerf98-536.
XX KW Mite allergen protein; map; high molecular weight; HW-map; allergy;
XX KW house dust mite; IgE; immunoglobulin E; allergen; mapA; mapB;
XX KW hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;
XX KW canine; veterinary; antibody; vaccine; immunisation.
XX OS Dermatophagoides farinae.
XX PN WO9954349-A2.
XX PD 28-OCT-1999.
XX PF 16-APR-1999; 99WO-US008524.
XX PR 17-APR-1998; 98US-00062013.
XX PR 13-MAY-1998; 98US-0085295P.
XX PR 02-SEP-1998; 98US-0098909P.
XX PA (HESK-) HESKA CORP.
XX PI Mccall CA, Hunter SW, Weber ER;
XX DR WPI; 2000-052700/04.
XX DR N-PSDB; AAZ38579, AAZ38580.

```

XX Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
PT used to modify an animals' hypersensitivity to mite allergens.  
XX  
XX Claim 3; Page 125-127; 154pp; English.  
XX  
XX This sequence represents Dermatophagoides farinae mite allergen protein  
CC (map) PDerf98-536, the mature form of PDerf98-555 (AAU96329). PDerf98-536  
CC has a molecular weight of 98 kD, comprising 536 amino acids, and is a  
CC component of the Dermatophagoides farinae high molecular weight mite  
CC allergen protein (HWM-map) composition. The HWM-map composition was  
CC isolated from a D. farinae homogenate by gel filtration, with each  
CC fraction being analysed for the presence of proteins that bound to IgE  
CC present in mite-allergic dog antisera. Mite allergenic proteins and  
CC peptides, and nucleic acids encoding them, may be used in therapeutic  
CC compositions to modify an animal's hypersensitivity reaction to mite  
CC allergens. Animals that may be treated include mammals and birds,  
CC especially felines, canines, equines, humans, other pets, and work or  
CC domestic animals. The proteins or fragments may also be used to diagnose  
CC allergies via a skin test. The proteins and peptides can also be used to  
CC raise antibodies, which have a variety of potential uses. For example,  
CC they can be used as vaccines to passively immunise animals against dust  
CC mite hypersensitivity, as positive controls in test kits and as tools to  
CC recover desired dust mite allergens from a mixture of proteins  
XX  
XX Sequence 536 AA;  
SQ

Query Match 94.3%; Score 99; DB 3; Length 536;  
Best Local Similarity 95.0%; Pred. No. 9.1e-07;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 DPAKGMSPPGFIVGEGVLS 20  
DB 279 DPAKGMSPPGFISGEGVLS 298

RESULT 11  
AAU96329  
XX AAU96329 standard; protein; 536 AA.  
XX  
XX 15-JUL-2002 (first entry)  
XX  
XX Der HWM-map polypeptide #16.  
XX  
XX Der HWM-map; American house dust mite; anti-allergic; mite; IgE;  
KW mite allergenic protein; immunoglobulin E; hypersensitivity;  
KW immunocomplex formation.  
XX  
XX Dermatophagoides farinae.  
OS  
XX WO200222807-A2.  
XX  
XX 21-MAR-2002.  
XX  
XX 14-SEP-2001; 2001WO-US028730.  
XX  
XX 14-SEP-2000; 2000US-00662293.  
XX  
XX (HESK-) HESKA CORP.  
PA  
XX Mccall CA, Hunter SW, Weber ER;  
XX  
XX WPI; 2002-351888/38.  
XX  
XX N-PSDB; ABK69575.  
XX  
XX New mite allergenic protein isolated from Dermatophagoides, designated  
PT Der HWM-map protein, useful as a vaccine for treating mite allergy.  
XX  
XX Claim 12; Page 125-127; 161pp; English.  
XX  
XX The invention relates to an isolated mite allergenic protein of

CC Dermatophagoides, designated Der HWM-map protein, and its related nucleic  
CC acid. The Der HWM-map protein is useful for eliciting an immune response  
CC against Der HWM-map protein. The protein or a reagent comprising a non-  
CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
CC cat) susceptible to or having an allergic response to a mite. A  
CC therapeutic composition is useful for desensitising a host animal to an  
CC allergic response to a mite. The DNA and protein can be used in the  
CC detection of anti-Der HWM-map antibodies in animal fluids, and inhibition  
CC of immunoglobulin (Ig) E or Der HWM-map protein activity associated with a  
CC disease. Antibodies that bind to Der HWM-map are useful for inhibiting  
CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
CC reducing hypersensitivity responses to mite allergens, and as vaccines  
CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
CC represent Der HWM-map polypeptides of the invention  
XX  
XX Sequence 536 AA;  
SQ

Query Match 94.3%; Score 99; DB 5; Length 536;  
Best Local Similarity 95.0%; Pred. No. 9.1e-07;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 DPAKGMSPPGFIVGEGVLS 20  
DB 279 DPAKGMSPPGFISGEGVLS 298

RESULT 12  
AAU96329  
XX AAU96329 standard; protein; 555 AA.  
XX  
XX 22-FEB-2000 (first entry)  
XX  
XX House dust mite (D. farinae) mite allergen protein (map) PDerf98-555.  
XX  
XX Mite allergen protein; map; high molecular weight; HWM-map; allergy;  
KW house dust mite; IgE; immunoglobulin E; allergen; mapA; mapB;  
KW hypersensitivity reaction; therapy; treatment; diagnosis; human; feline;  
KW canine; veterinary; antibody; vaccine; immunisation.  
XX  
XX Dermatophagoides farinae.  
OS  
XX  
XX Key Location/Qualifiers  
FH Peptide 1..19  
FT /note= "Signal peptide"  
FT Protein 20..555  
FT /note= "Mature PDerf98-555"  
XX  
XX WO9954349-A2.  
XX  
XX 28-OCT-1999.  
XX  
XX 16-APR-1999; 99WO-US008524.  
XX  
XX 17-APR-1998; 98US-00062013.  
XX  
XX 13-MAY-1998; 98US-0085295P.  
XX  
XX 02-SEP-1998; 98US-0098909P.  
XX  
XX (HESK-) HESKA CORP.  
XX  
XX Mccall CA, Hunter SW, Weber ER;  
XX  
XX WPI; 2000-052700/04.  
XX  
XX N-PSDB; AAU96329, AAU963576, AAU963577, AAU963578.  
XX  
XX Novel high molecular weight Dermatophagoides nucleic acid polypeptides  
PT used to modify an animals' hypersensitivity to mite allergens.  
XX  
XX Claim 3; Page 111-113; 154pp; English.  
XX  
XX This sequence represents Dermatophagoides farinae mite allergen protein  
CC (map) PDerf98-555. PDerf98-555 has a molecular weight of 98 kD,

CC comprising 555 amino acids, and is a component of the Dermatophagoides  
 CC farinae high molecular weight mite allergen protein (HWM-map)  
 CC composition. The HWM-map composition was isolated from a D. farinae  
 CC homogenate by gel filtration, with each fraction being analysed for the  
 CC presence of proteins that bound to IgE present in mite-allergic dog  
 CC antisera. Mite allergenic proteins and peptides, and nucleic acids  
 CC encoding them, may be used in therapeutic compositions to modify an  
 CC animal's hypersensitivity reaction to mite allergens. Animals that may be  
 CC treated include mammals and birds, especially felines, canines, equines,  
 CC humans, other pets, and work or domestic animals. The proteins or  
 CC fragments may also be used to diagnose allergies via a skin test. The  
 CC proteins and peptides can also be used to raise antibodies, which have a  
 CC variety of potential uses. For example, they can be used as vaccines to  
 CC passively immunise animals against dust mite hypersensitivity, as  
 CC positive controls in test kits and as tools to recover desired dust mite  
 CC allergens from a mixture of proteins  
 XX  
 SQ Sequence 555 AA;

Query Match 94.3%; Score 99; DB 3; Length 555;  
 Best Local Similarity 95.0%; Pred. No. 9.5e-07;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DPAKGMSPPGFIVGEGVLS 20  
 DB 298 DPAKGMSPPGFISGEGVLS 317  
 |||||

RESULT 13  
 AAU96327  
 ID AAU96327 standard; protein; 555 AA.

XX AAU96327;  
 DT 15-JUL-2002 (first entry)  
 DE Der HWM-map polypeptide #14.  
 KW Der HWM-map; American house dust mite; antiallergic; mite; IgE;  
 KW mite allergenic protein; immunoglobulin E; hypersensitivity;  
 KW immunocomplex formation.  
 XX Dermatophagoides farinae.  
 XX WO200222807-A2.

XX 21-MAR-2002.  
 PF 14-SEP-2001; 2001WO-US028730.  
 PR 14-SEP-2000; 2000US-00662293.  
 XX (HESK-) HESKA CORP.  
 XX McCall CA, Hunter SW, Weber ER;  
 XX WPI; 2002-351888/38.  
 DR N-PSDB; ABK69571.

PT New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HWM-map protein, useful as a vaccine for treating mite allergy.  
 XX  
 PS Claim 12; Page 114-116; 161pp; English.

XX The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HWM-map protein, and its related nucleic  
 CC acid. The Der HWM-map protein is useful for eliciting an immune response  
 CC against Der HWM-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitising a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HWM-map antibodies in animal fluids, and inhibition

CC of immunoglobulin (Ig)E or Der HWM-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HWM-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HWM-map polypeptides of the invention  
 XX  
 SQ Sequence 555 AA;

Query Match 94.3%; Score 99; DB 5; Length 555;  
 Best Local Similarity 95.0%; Pred. No. 9.5e-07;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DPAKGMSPPGFIVGEGVLS 20  
 DB 298 DPAKGMSPPGFISGEGVLS 317  
 |||||

RESULT 14  
 AAU96328  
 ID AAU96328 standard; protein; 555 AA.

XX AAU96328;  
 AC AAU96328;  
 DT 15-JUL-2002 (first entry)  
 DE Der HWM-map polypeptide #15.  
 KW Der HWM-map; American house dust mite; antiallergic; mite; IgE;  
 KW mite allergenic protein; immunoglobulin E; hypersensitivity;  
 KW immunocomplex formation.  
 XX Dermatophagoides farinae.  
 XX WO200222807-A2.

XX 21-MAR-2002.  
 PF 14-SEP-2001; 2001WO-US028730.  
 PR 14-SEP-2000; 2000US-00662293.

XX (HESK-) HESKA CORP.  
 XX McCall CA, Hunter SW, Weber ER;  
 XX WPI; 2002-351888/38.  
 DR N-PSDB; ABK69573.

PT New mite allergenic protein isolated from Dermatophagoides, designated  
 PT Der HWM-map protein, useful as a vaccine for treating mite allergy.  
 XX  
 PS Claim 12; Page 120-122; 161pp; English.

XX The invention relates to an isolated mite allergenic protein of  
 CC Dermatophagoides, designated Der HWM-map protein, and its related nucleic  
 CC acid. The Der HWM-map protein is useful for eliciting an immune response  
 CC against Der HWM-map protein. The protein or a reagent comprising a non-  
 CC proteinaceous epitope is useful for identifying an animal (e.g., dog,  
 CC cat) susceptible to or having an allergic response to a mite. A  
 CC therapeutic composition is useful for desensitising a host animal to an  
 CC allergic response to a mite. The DNA and protein can be used in the  
 CC detection of anti-Der HWM-map antibodies in animal fluids, and inhibition  
 CC of immunoglobulin (Ig)E or Der HWM-map protein activity associated with a  
 CC disease. Antibodies that bind to Der HWM-map are useful for inhibiting  
 CC binding of proteins to IgE, to prevent immunocomplex formation, thus  
 CC reducing hypersensitivity responses to mite allergens, and as vaccines  
 CC against mite allergen hypersensitivity. Sequences AAU96314-AAU96342  
 CC represent Der HWM-map polypeptides of the invention  
 XX  
 SQ Sequence 555 AA;

Query Match 94.3%; Score 99; DB 5; Length 555;

Best Local Similarity 95.0%; Pred. No. 9.5e-07;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DPAKGMSPGFIYVGEVLS 20  
Db 298 DPAKGMSPGFIYVGEVLS 317

RESULT 15  
ABB68081  
ID ABB68081 standard; protein; 488 AA.  
XX AC ABB68081;  
XX DT 26-MAR-2002 (first entry)  
XX DE Drosophila melanogaster polypeptide SEQ ID NO 31035.  
XX KW Drosophila; developmental biology; cell signalling; insecticide;  
XX KW pharmaceutical.  
XX OS Drosophila melanogaster.  
XX WO200171042-A2.  
XX PD 27-SEP-2001.  
XX PF 23-MAR-2001; 2001WO-US009231.  
XX PR 23-MAR-2000; 2000US-0191637P.  
XX ER 11-JUL-2000; 2000US-00614150.  
XX FA (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
XX DR N-PSDB; ABL12184.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX Disclosure; SEQ ID NO 31035; 21pp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
XX ABB72072). The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pt\_sequences

SQ Sequence 488 AA;  
Query Match 47.6%; Score 50; DB 4; Length 488;  
Best Local Similarity 40.0%; Pred. No. 33;  
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 DPAKGMSPGFIYVGEVLS 20  
Db 407 DPQSGFDPFTYLGDELS 426

Search completed: March 22, 2004, 06:51:43  
Job time : 6.08116 secs

